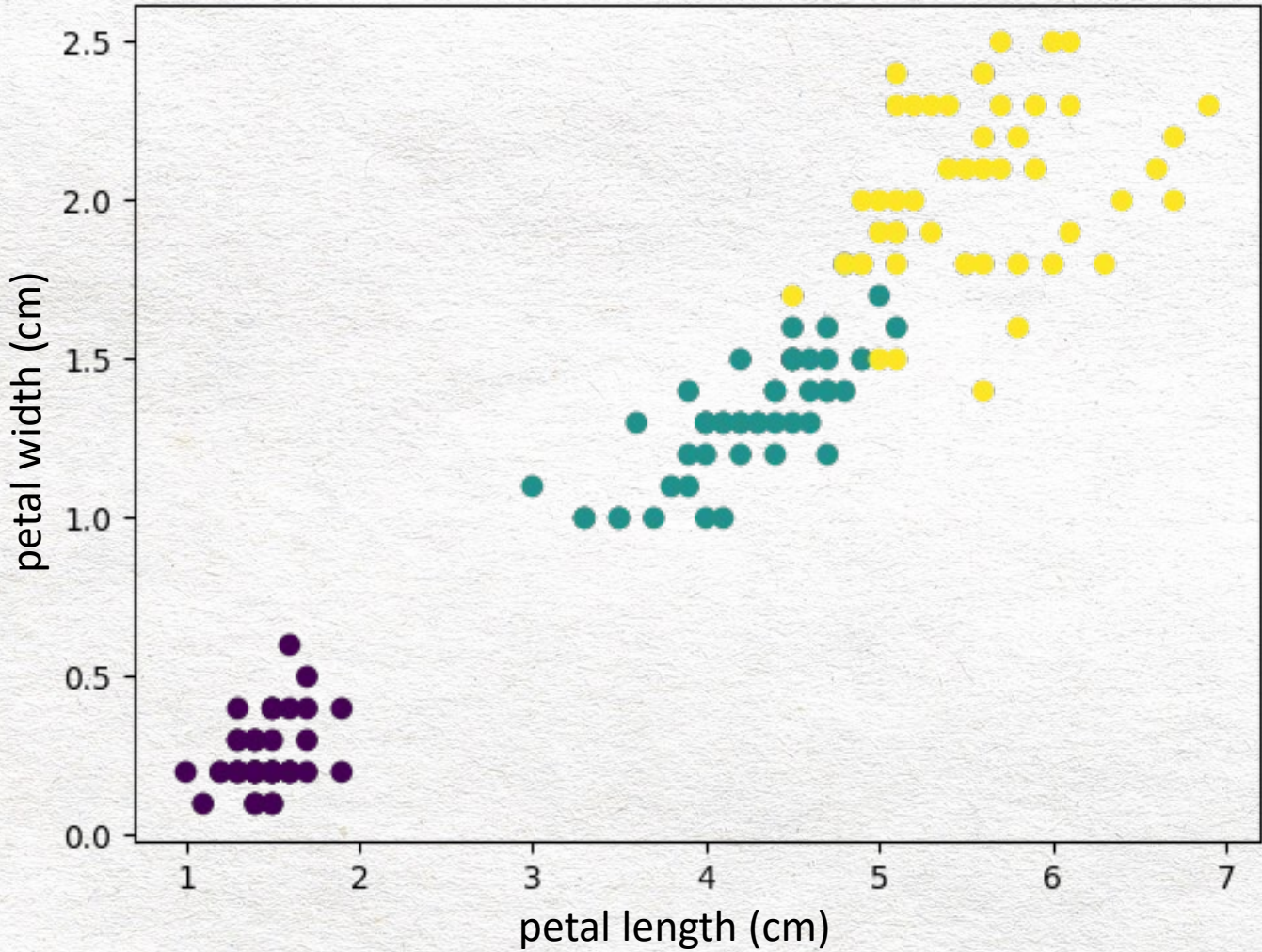


PCA (Principal Component Analysis)

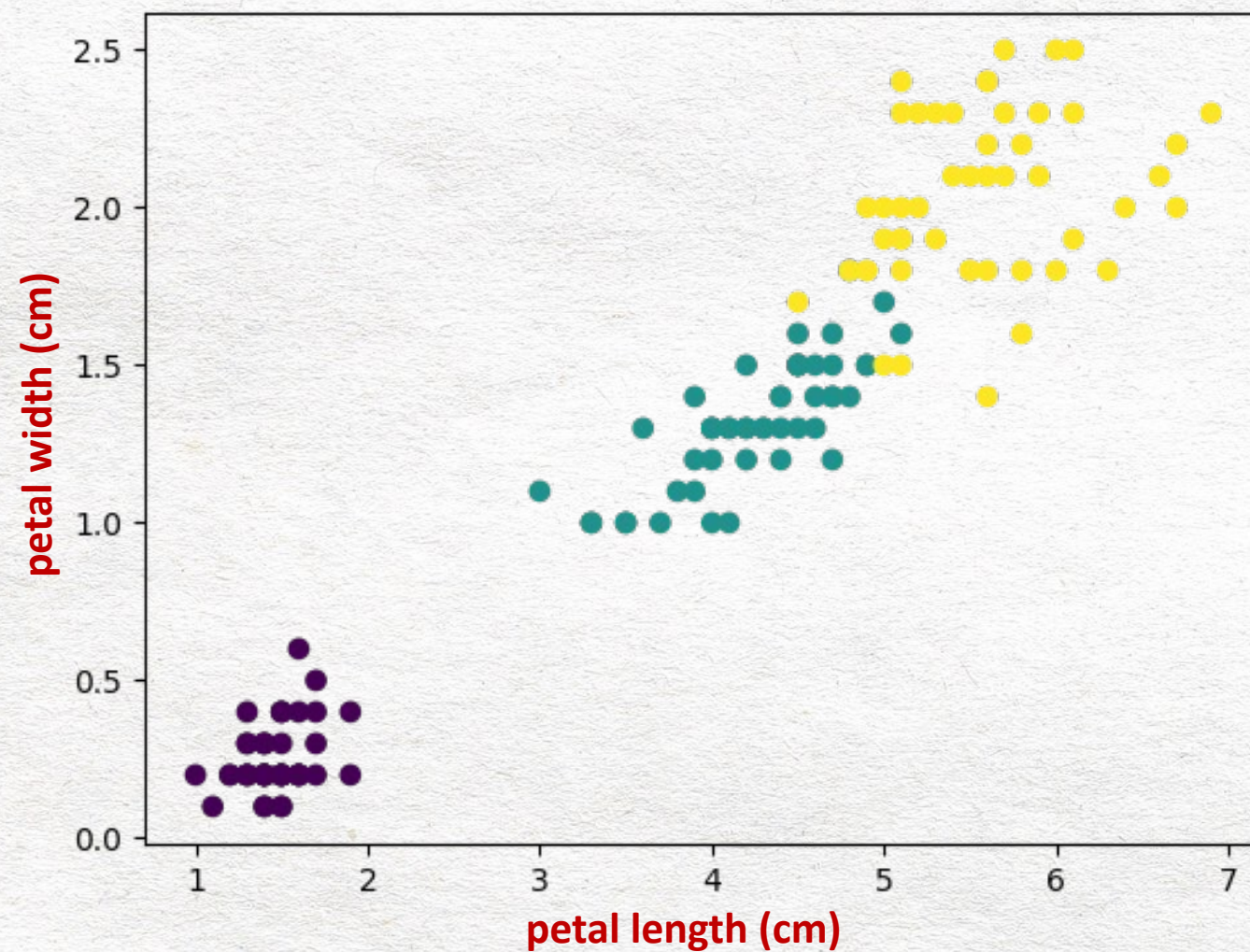
Quando representamos o dataset iris utilizando apenas 2 colunas, **não estamos considerando todas as informações desse dataset**



	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

PCA (Principal Component Analysis)

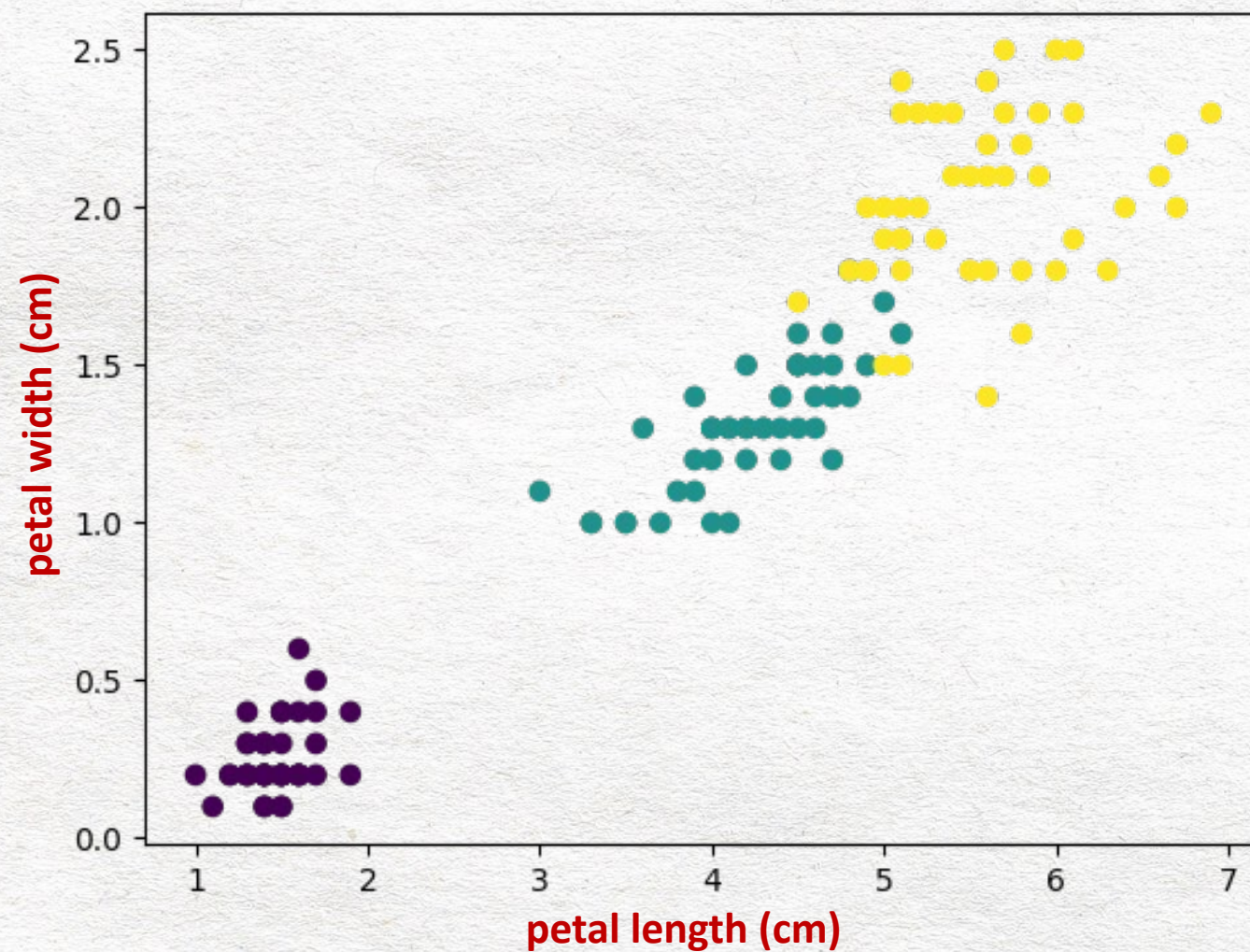
Quando representamos o dataset iris utilizando apenas 2 colunas, **não estamos considerando todas as informações desse dataset**



	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

PCA (Principal Component Analysis)

Quando representamos o dataset iris utilizando apenas 2 colunas, **não estamos considerando todas as informações desse dataset**

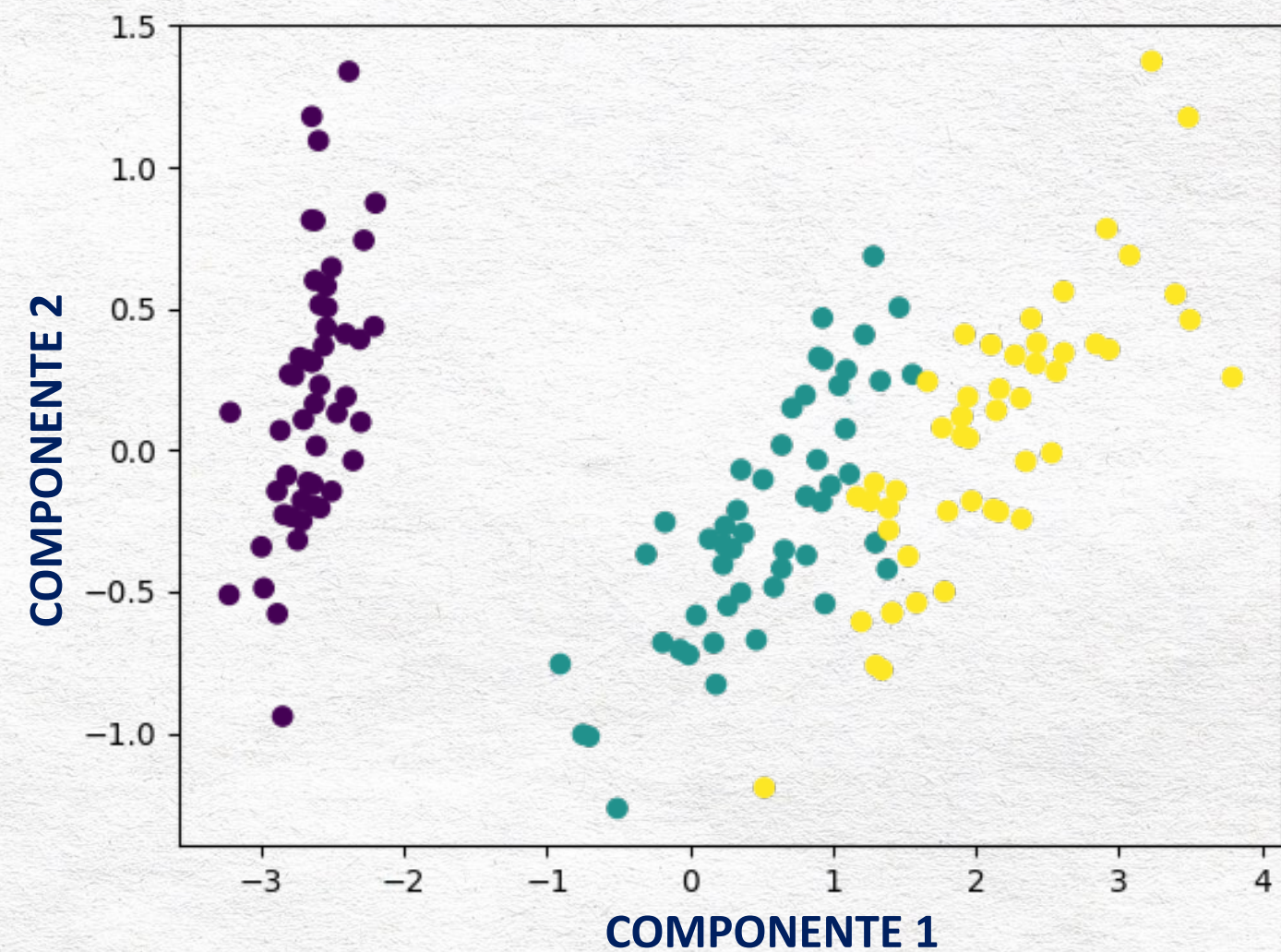
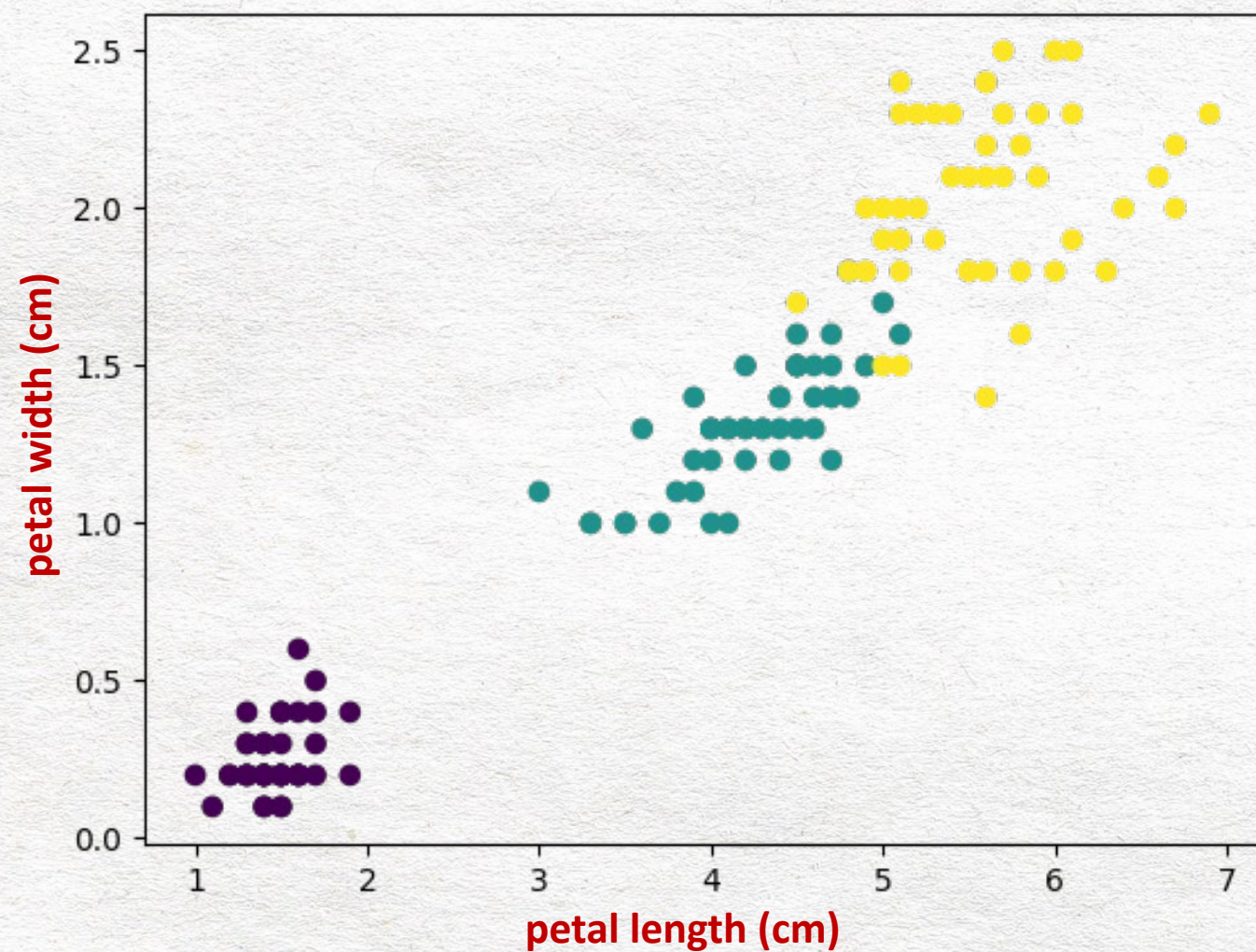


	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
...
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

E essas colunas?

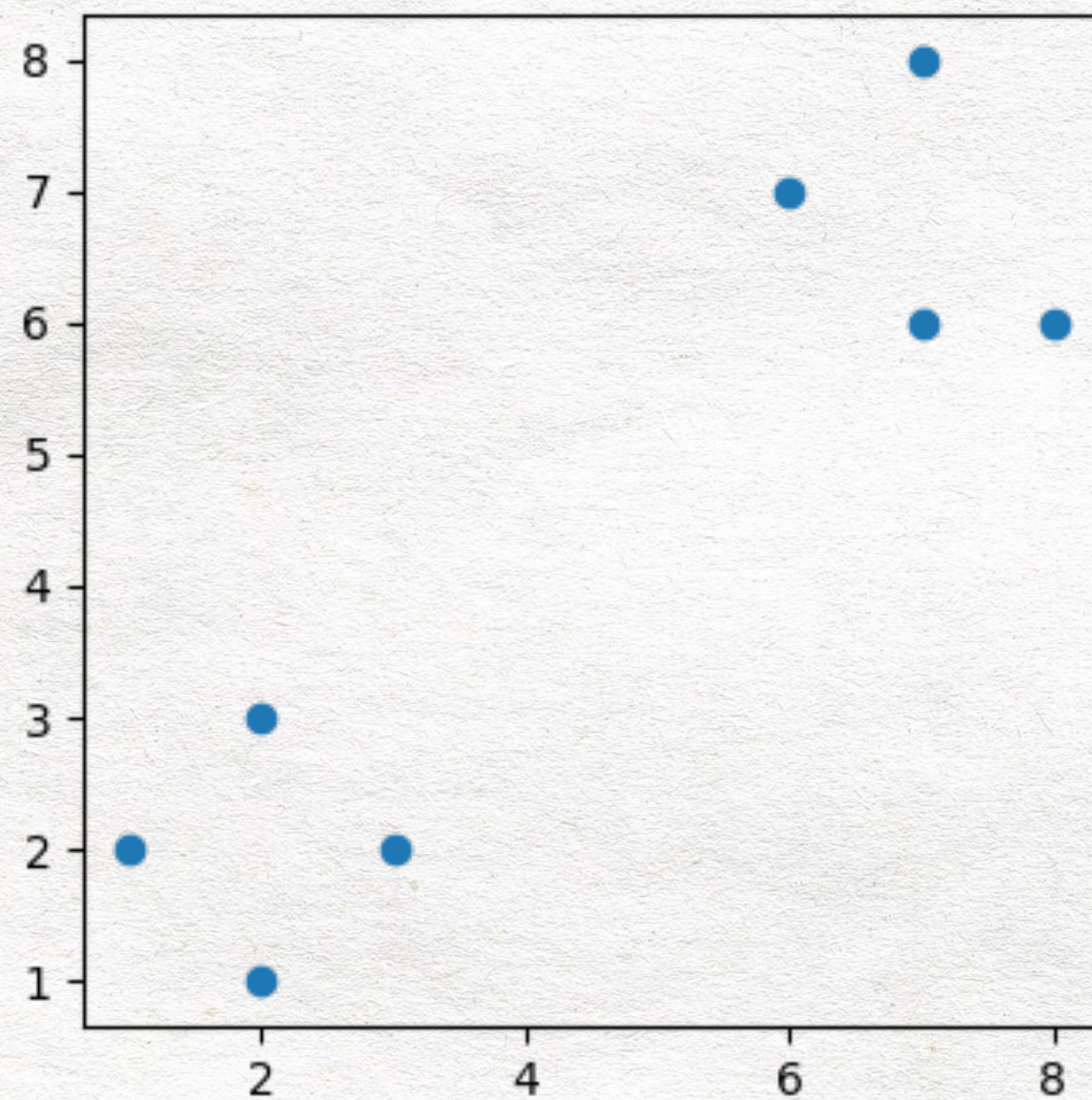
PCA (Principal Component Analysis)

O PCA vai fazer a redução linear da dimensionalidade dos dados, de forma a **permitir sua visualização em 2D perdendo menos informações**



PCA (Principal Component Analysis)

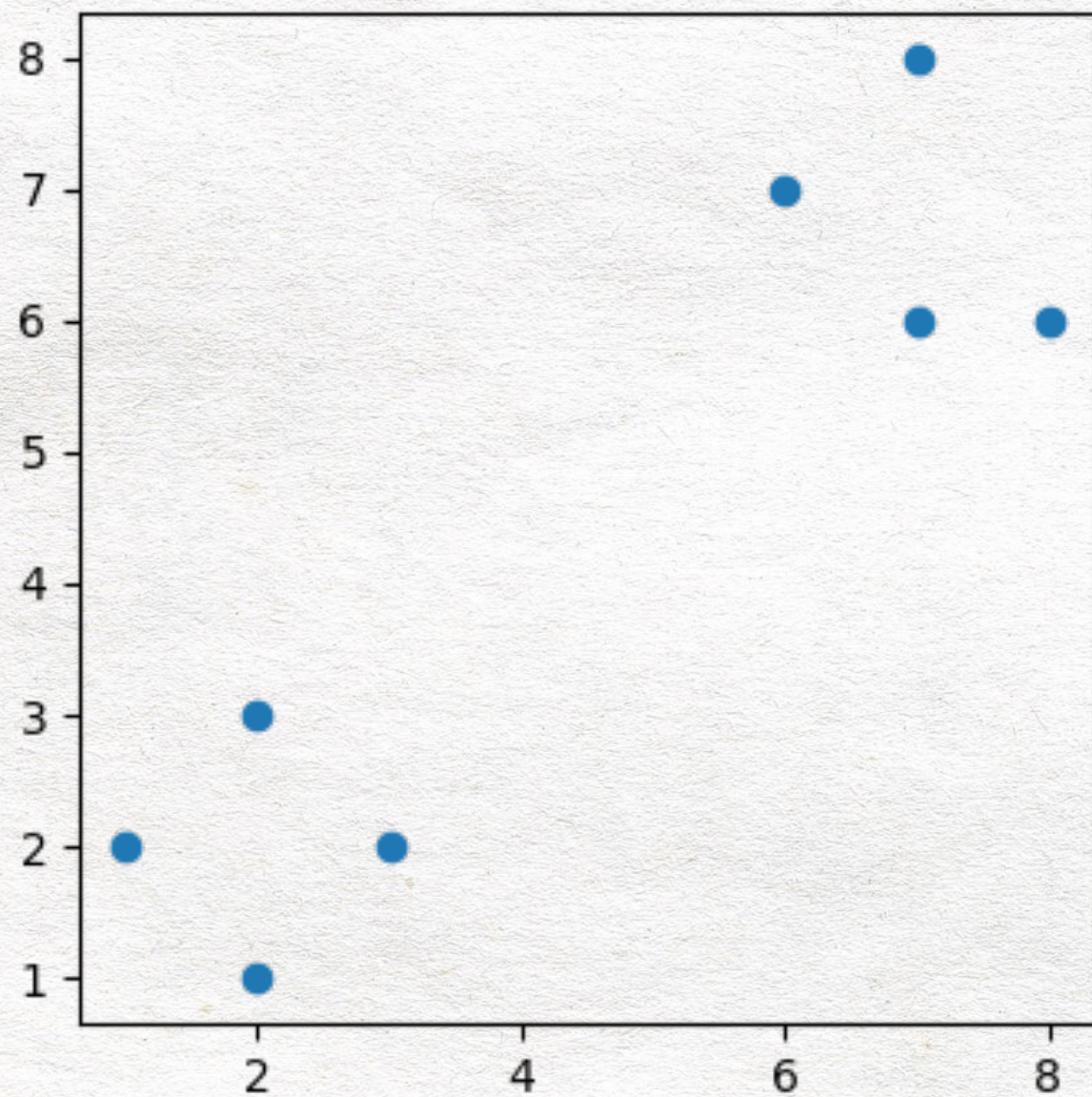
Como vamos fazer o PCA?



	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

PCA (Principal Component Analysis)

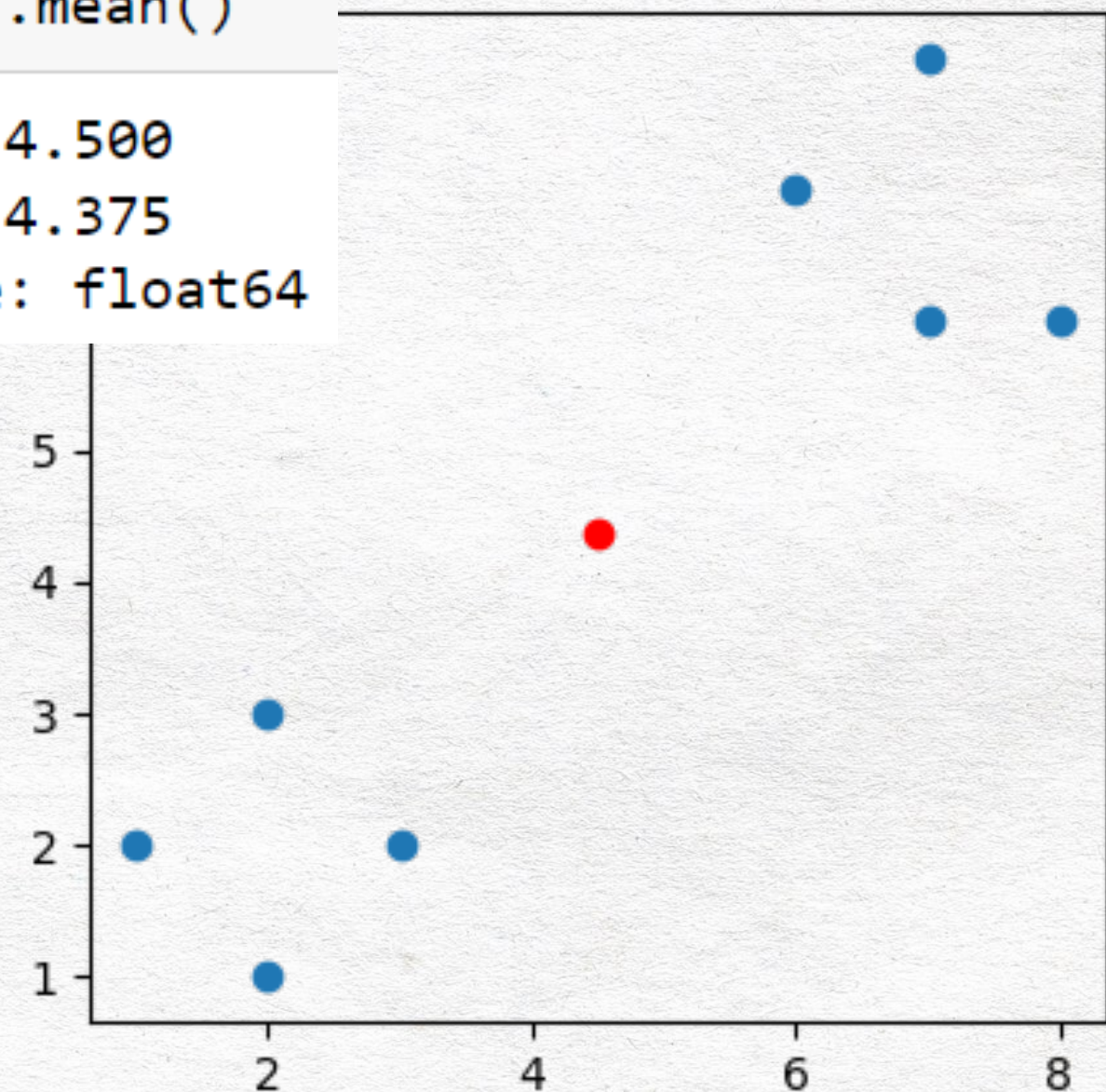
Como vamos fazer o PCA?



	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

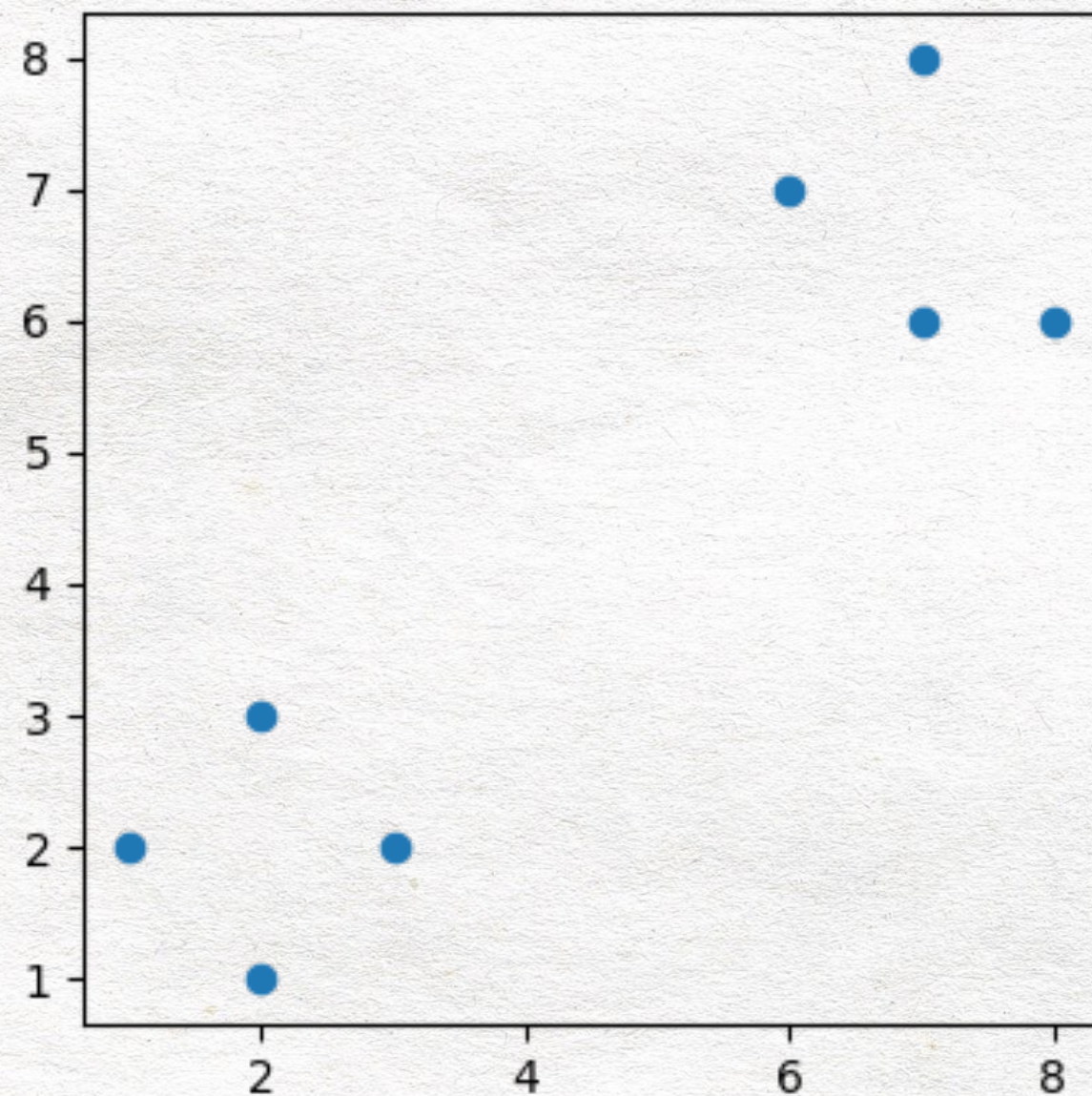
O primeiro passo é tirar a média de X e Y

```
dados.mean()  
x    4.500  
y    4.375  
dtype: float64
```



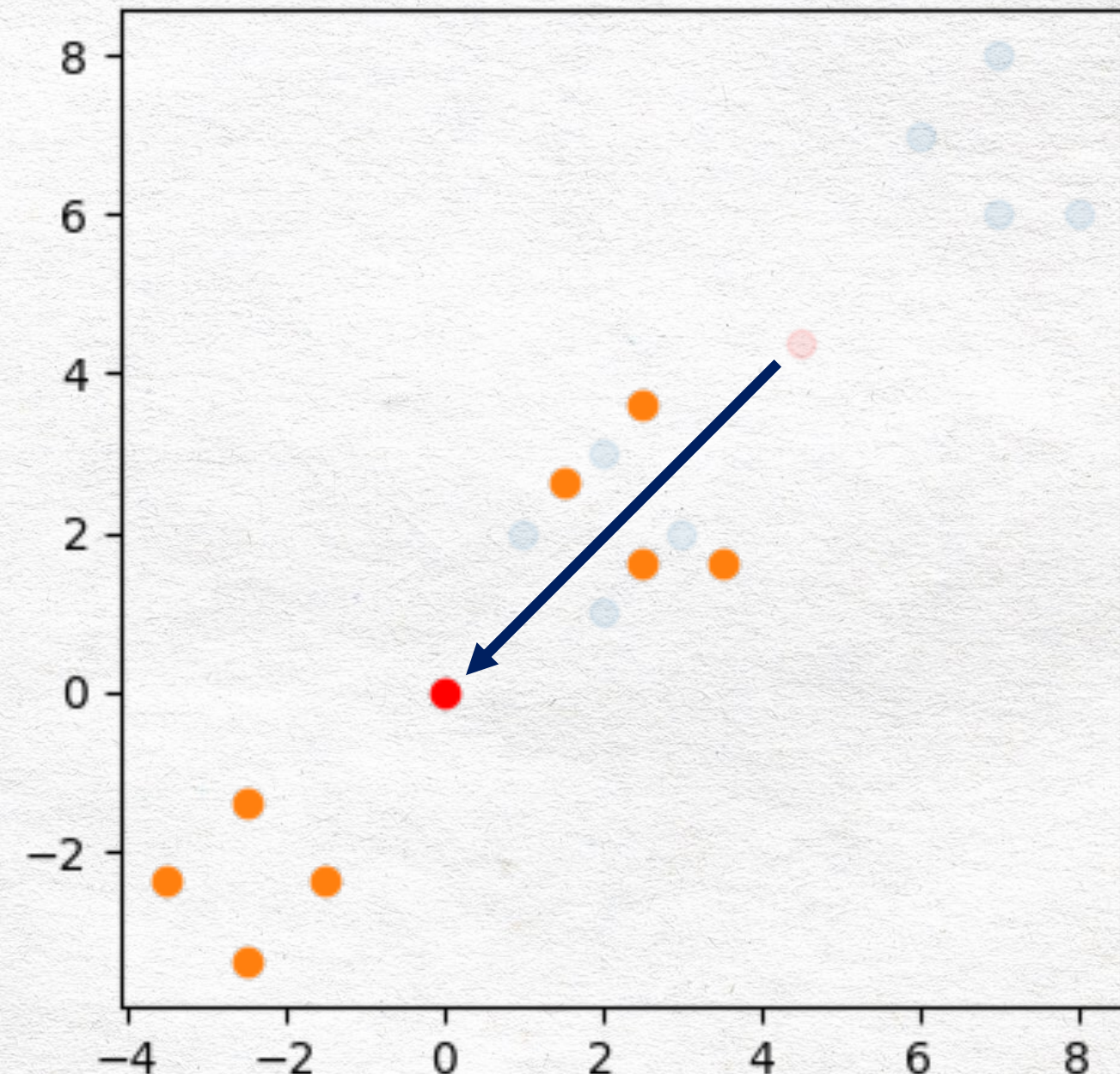
PCA (Principal Component Analysis)

Como vamos fazer o PCA?



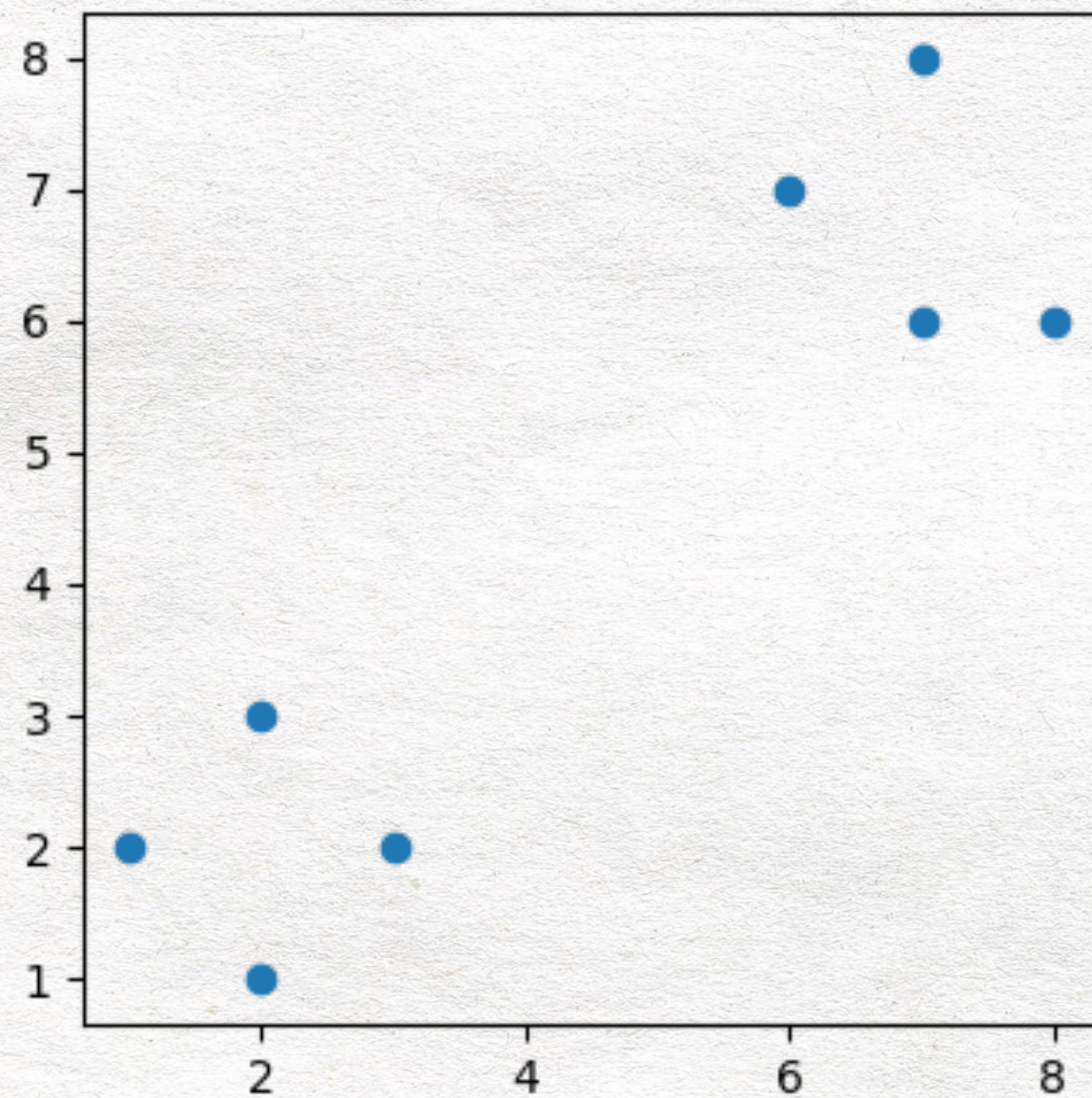
	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

Em seguida, vamos deslocar todos os dados até o centro deles (a média em vermelho) chegar na origem (0,0) do gráfico



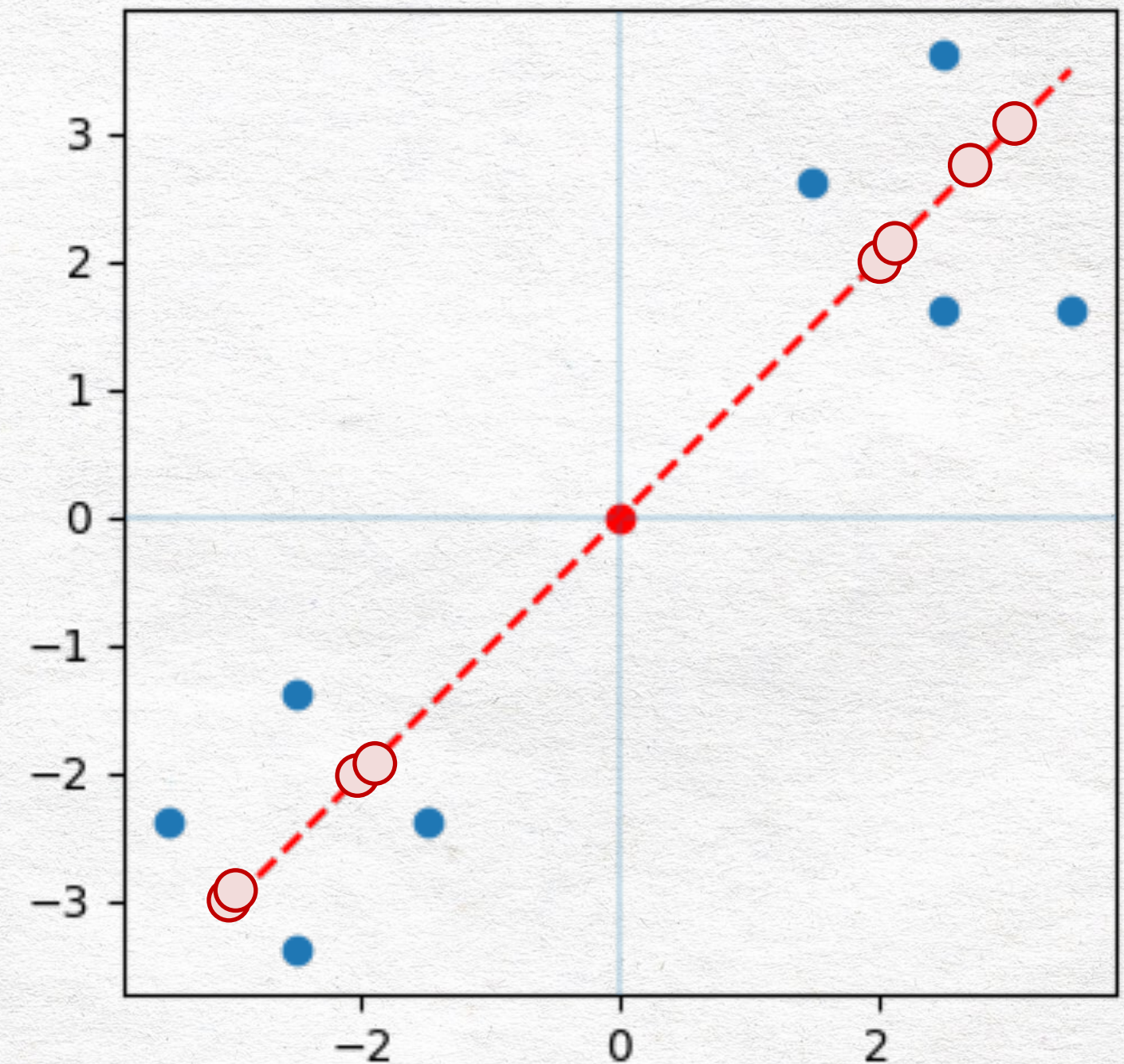
PCA (Principal Component Analysis)

Como vamos fazer o PCA?



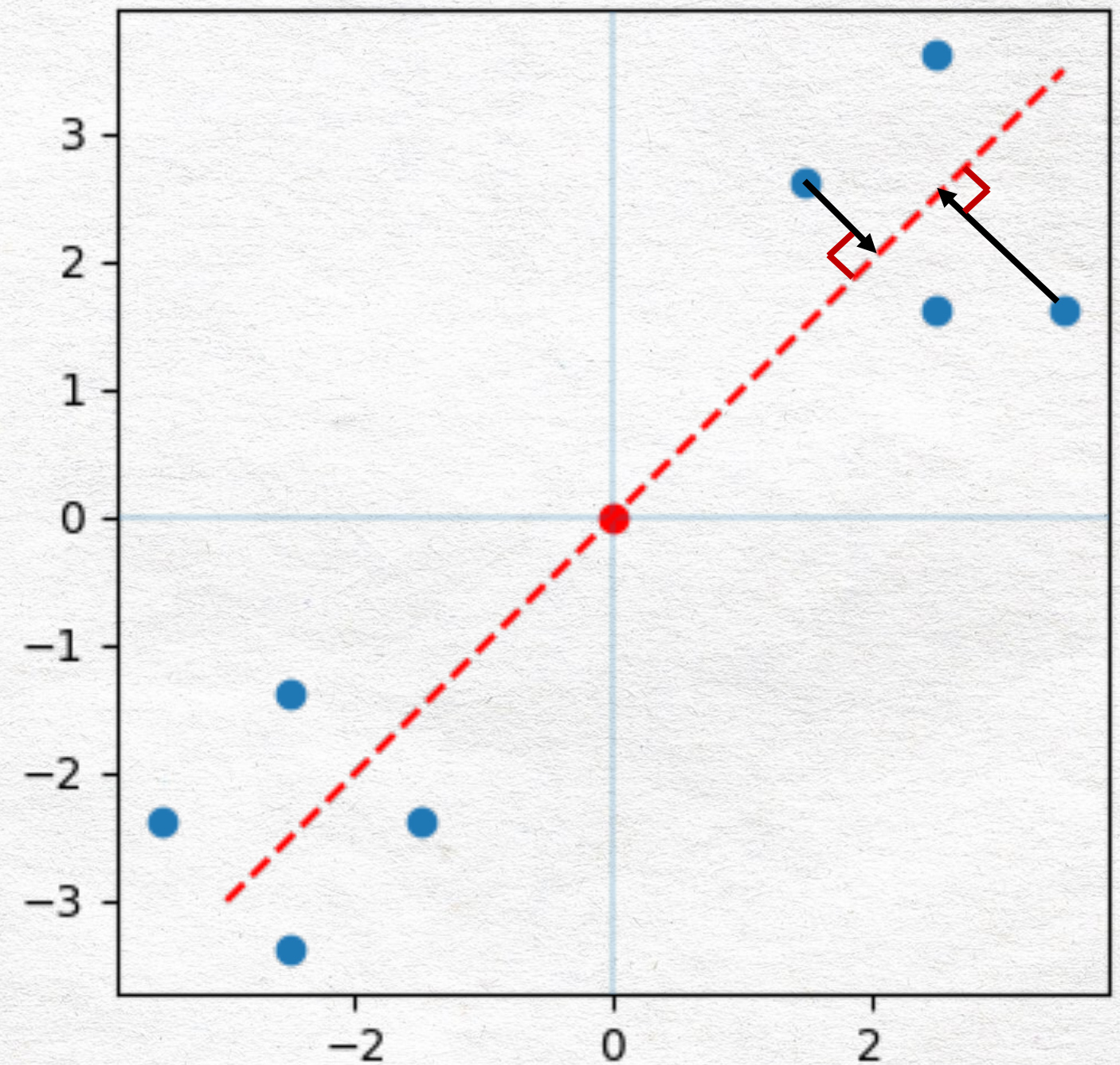
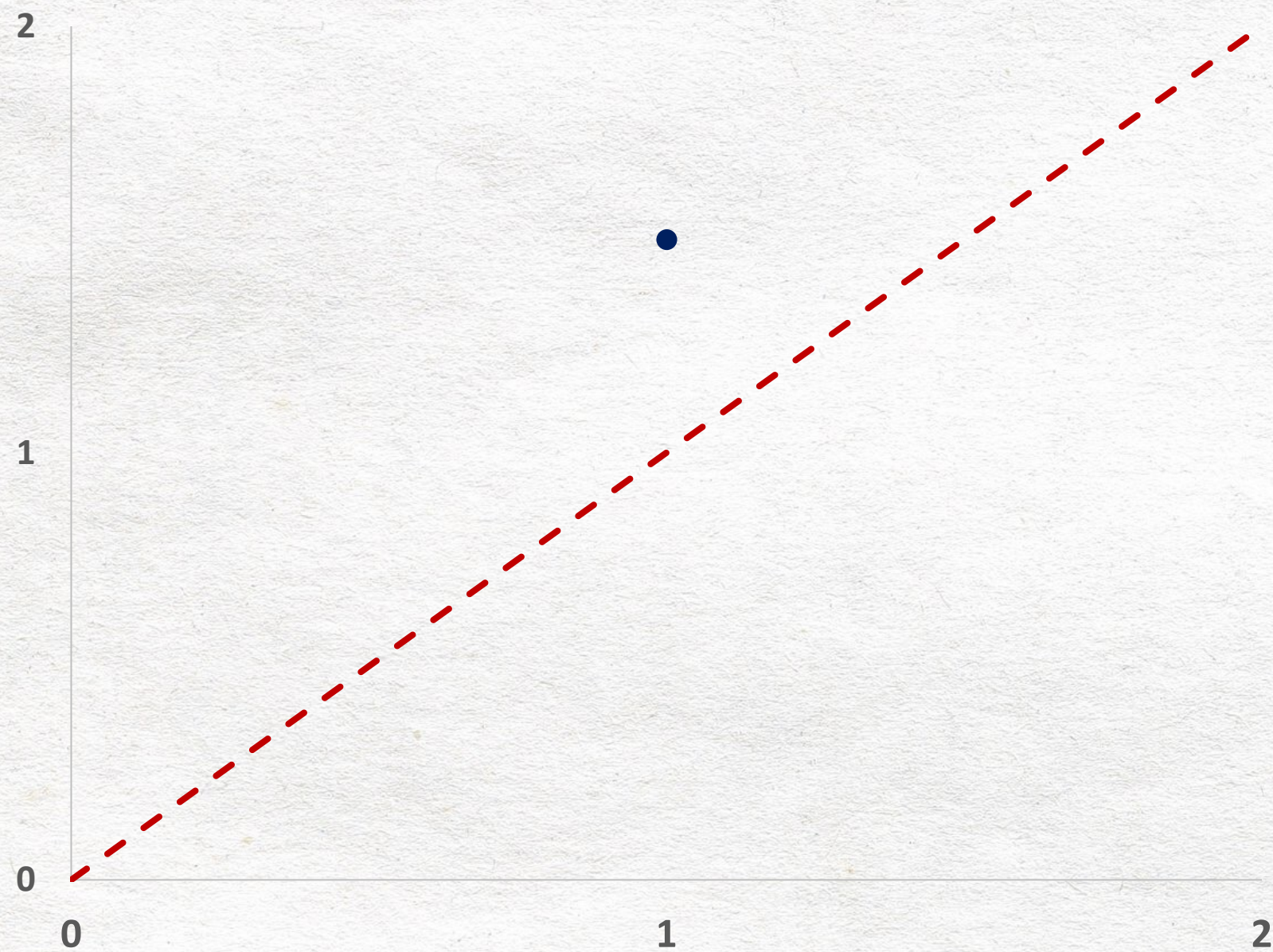
	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

E vamos buscar a reta que passa pelo centro (0,0) e **minimiza a distância dos pontos a sua projeção na reta**



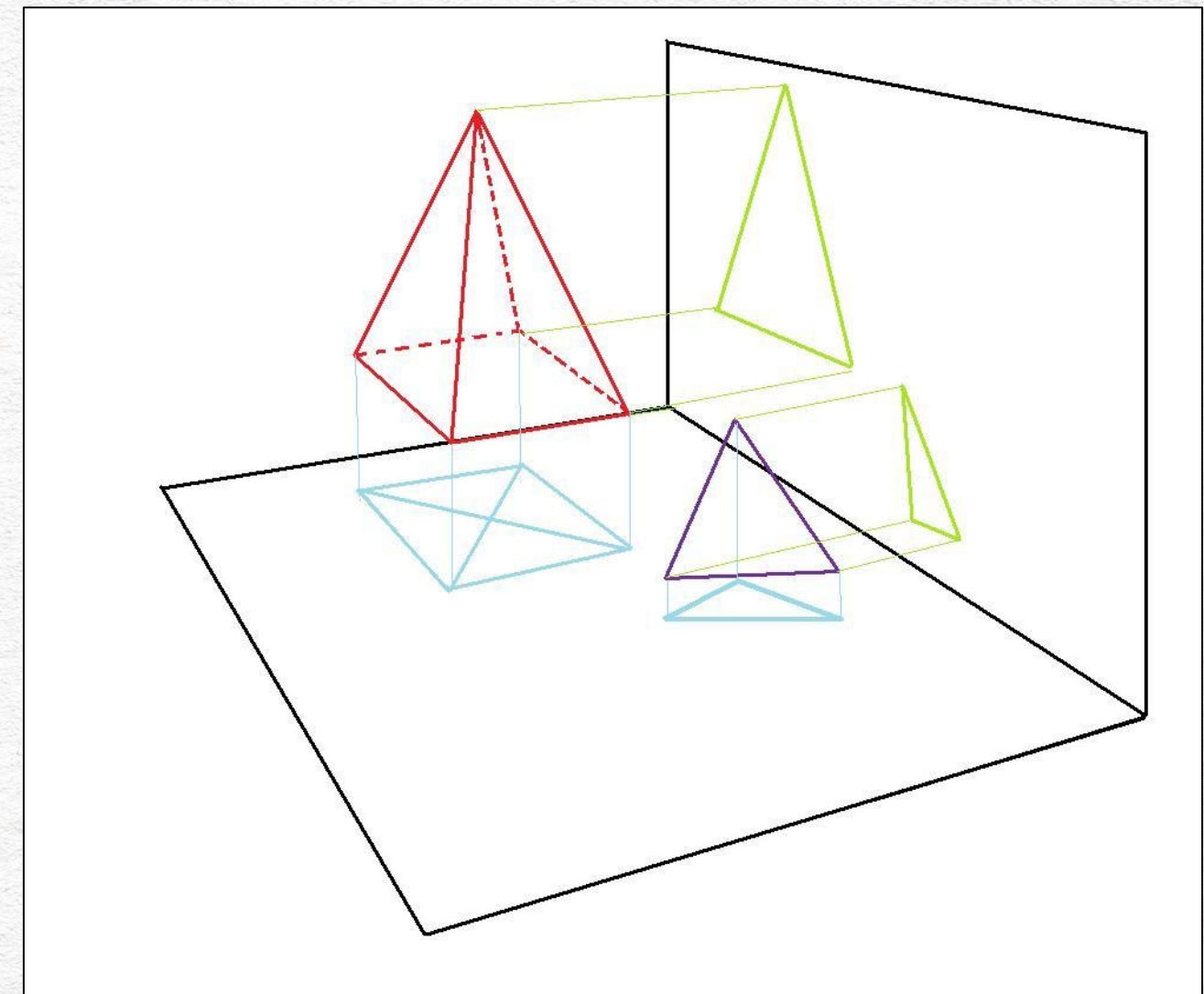
PCA (Principal Component Analysis)

O que é uma projeção?



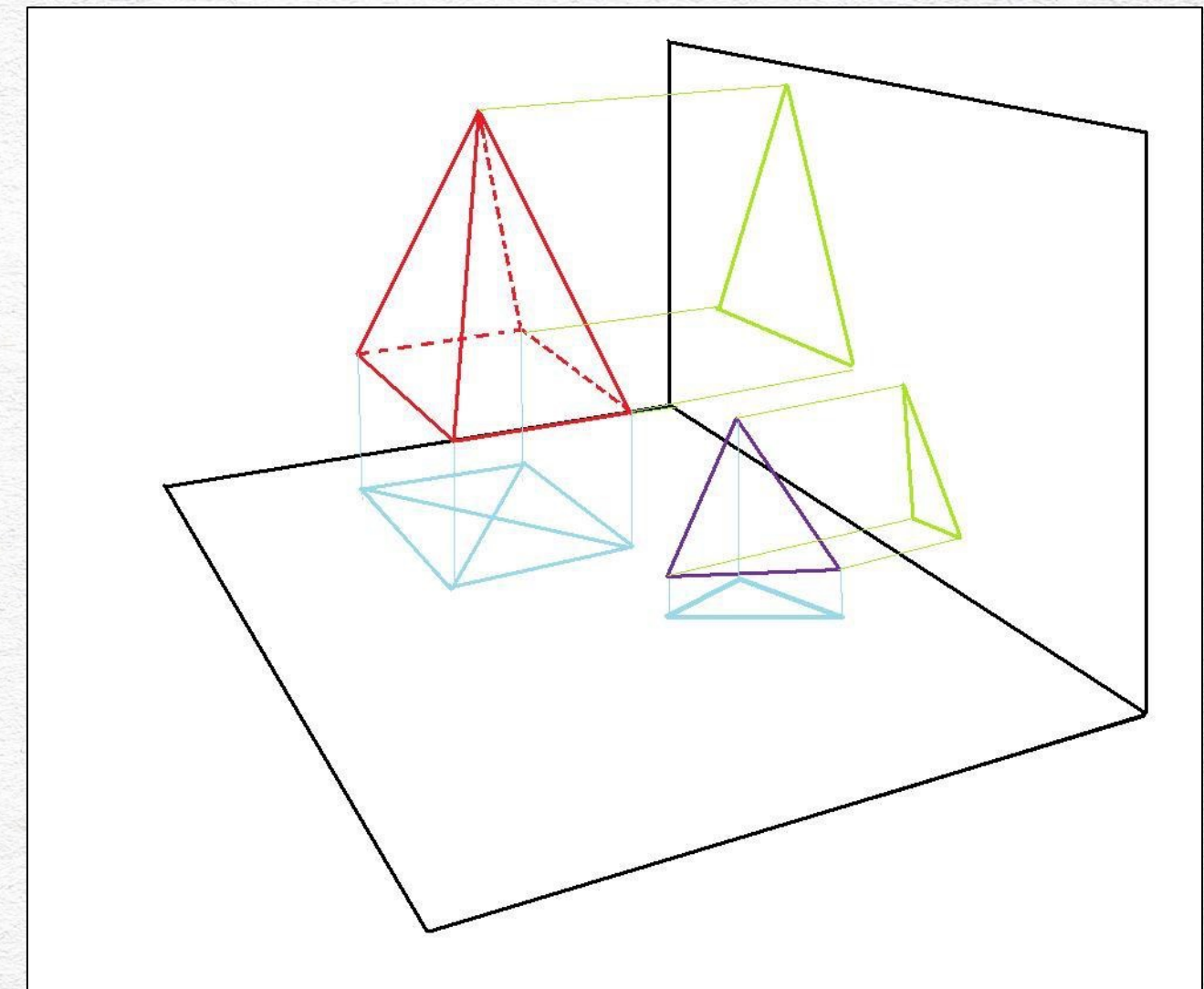
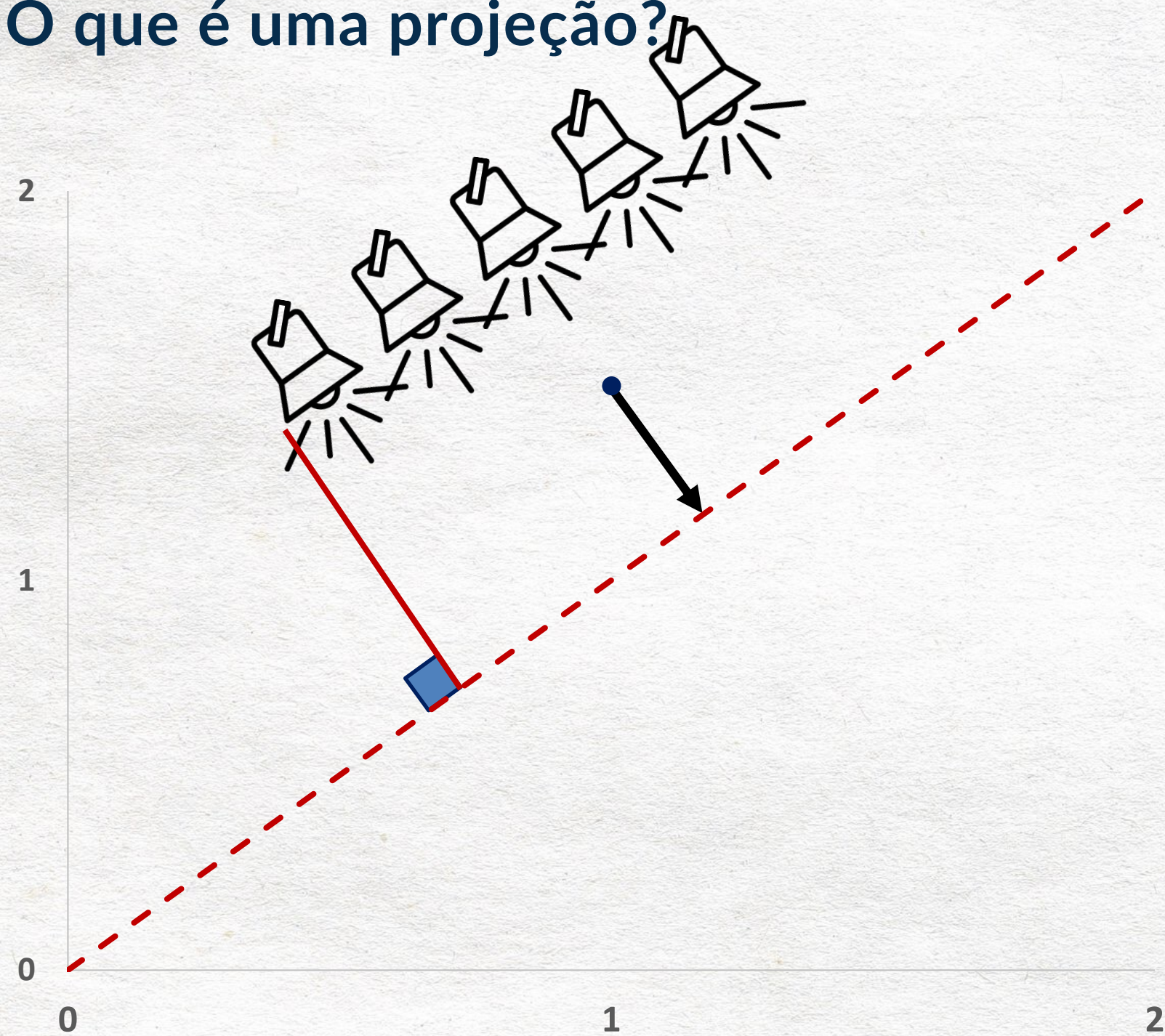
PCA (Principal Component Analysis)

O que é uma projeção?



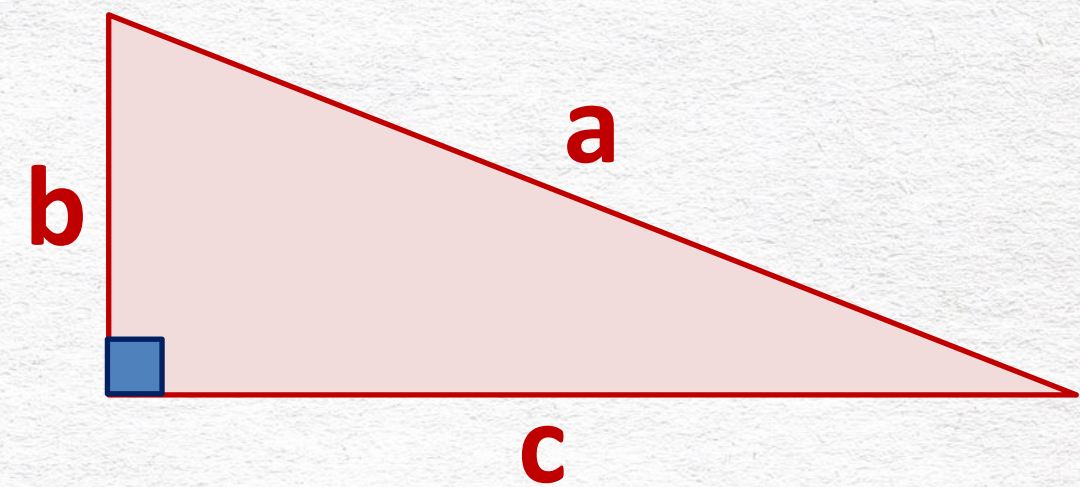
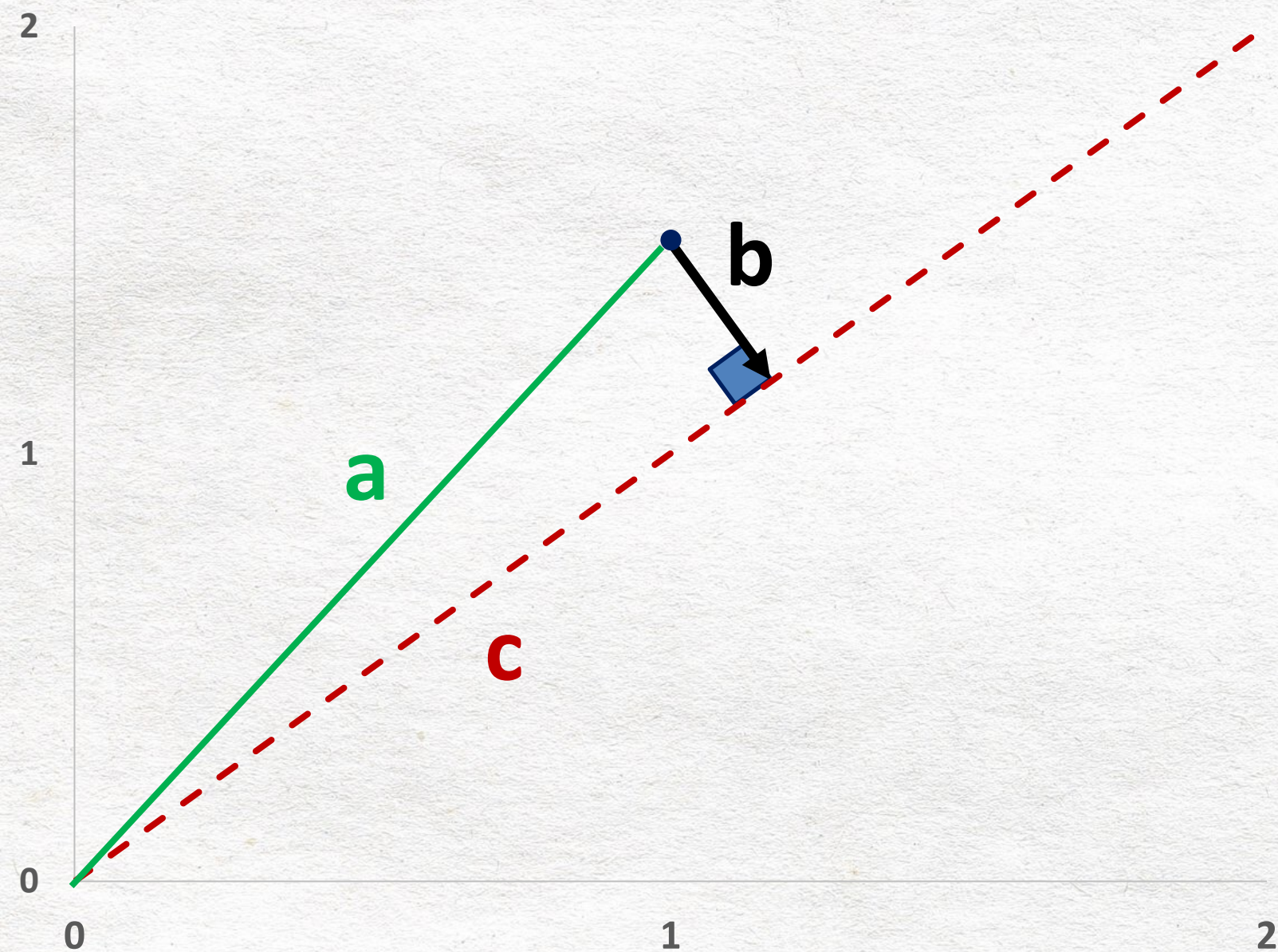
PCA (Principal Component Analysis)

O que é uma projeção?



PCA (Principal Component Analysis)

O que é uma projeção?



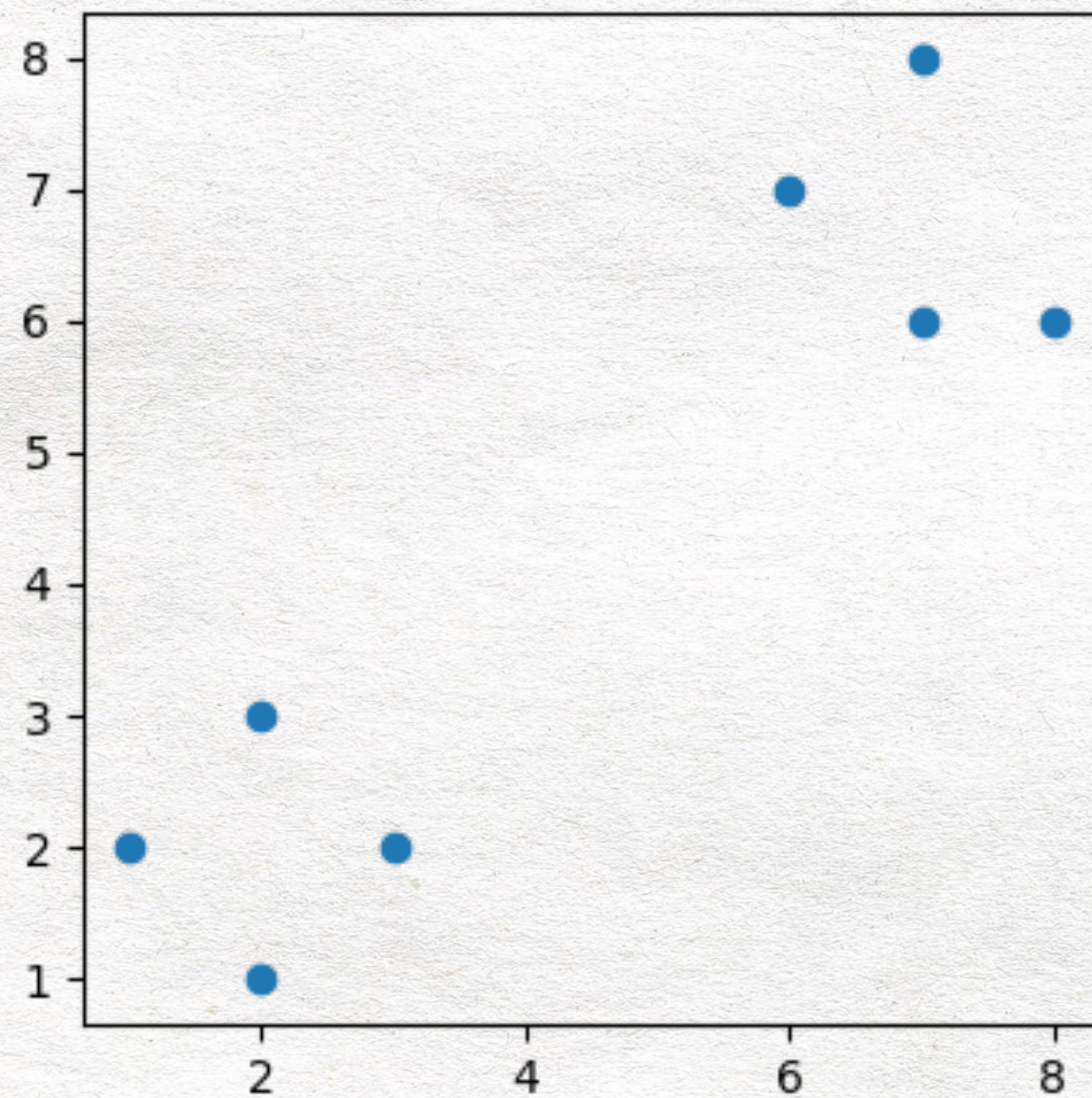
Teorema de Pitágoras

$$a^2 = b^2 + c^2$$

(para um triângulo retângulo)

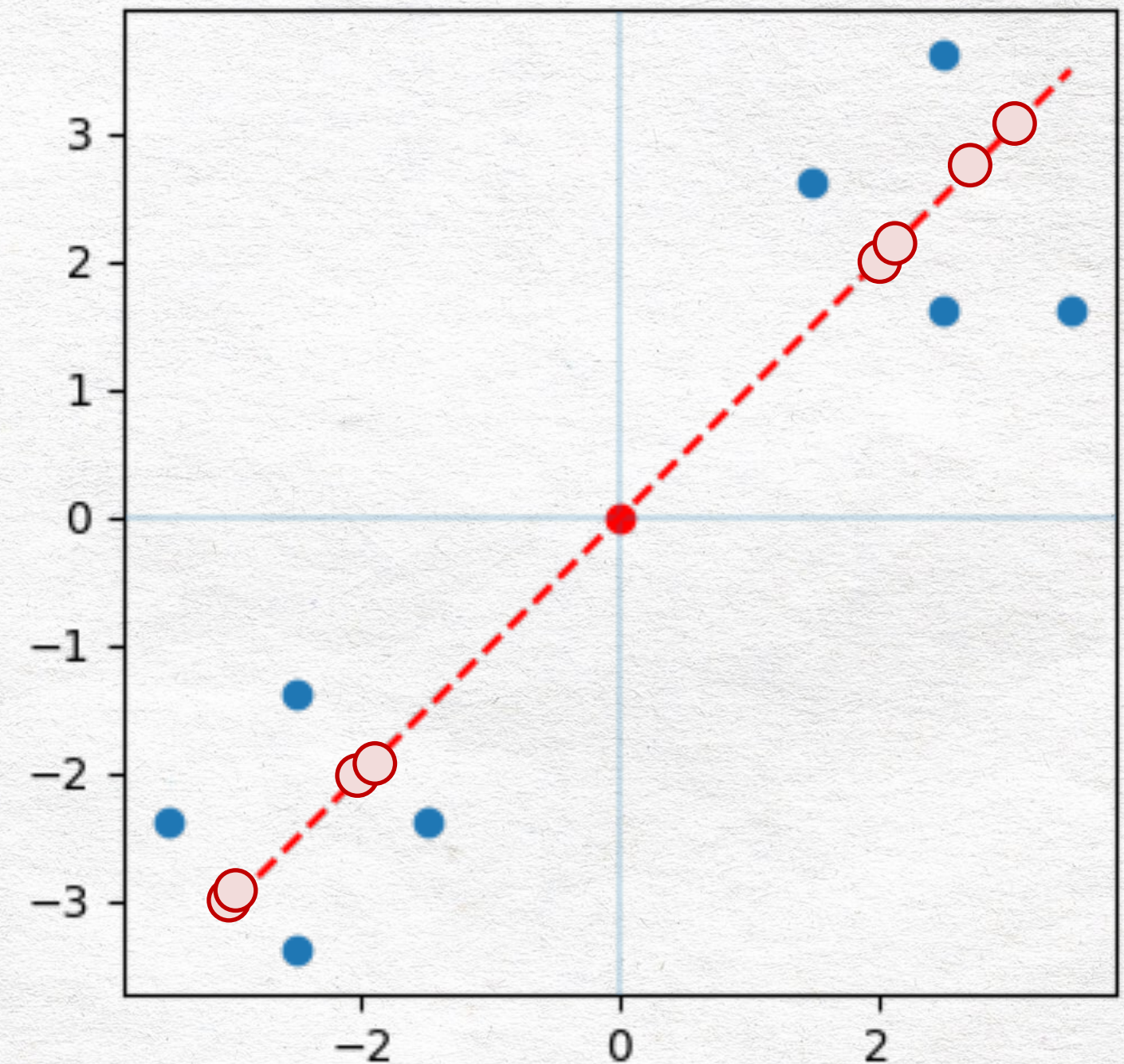
PCA (Principal Component Analysis)

Como vamos fazer o PCA?



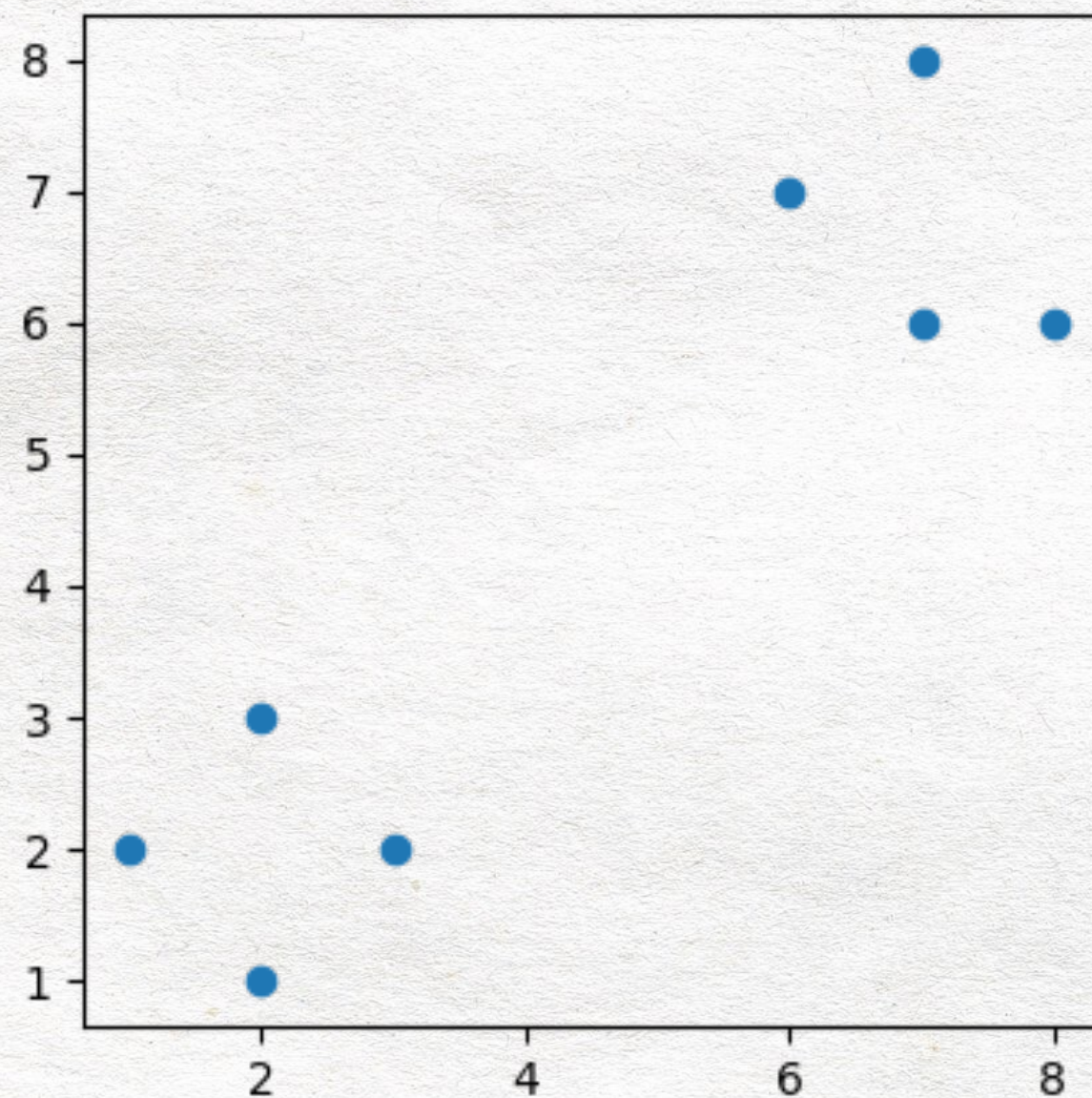
	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

E vamos buscar a reta que passa pelo centro (0,0) e **minimiza a distância dos pontos a sua projeção na reta**



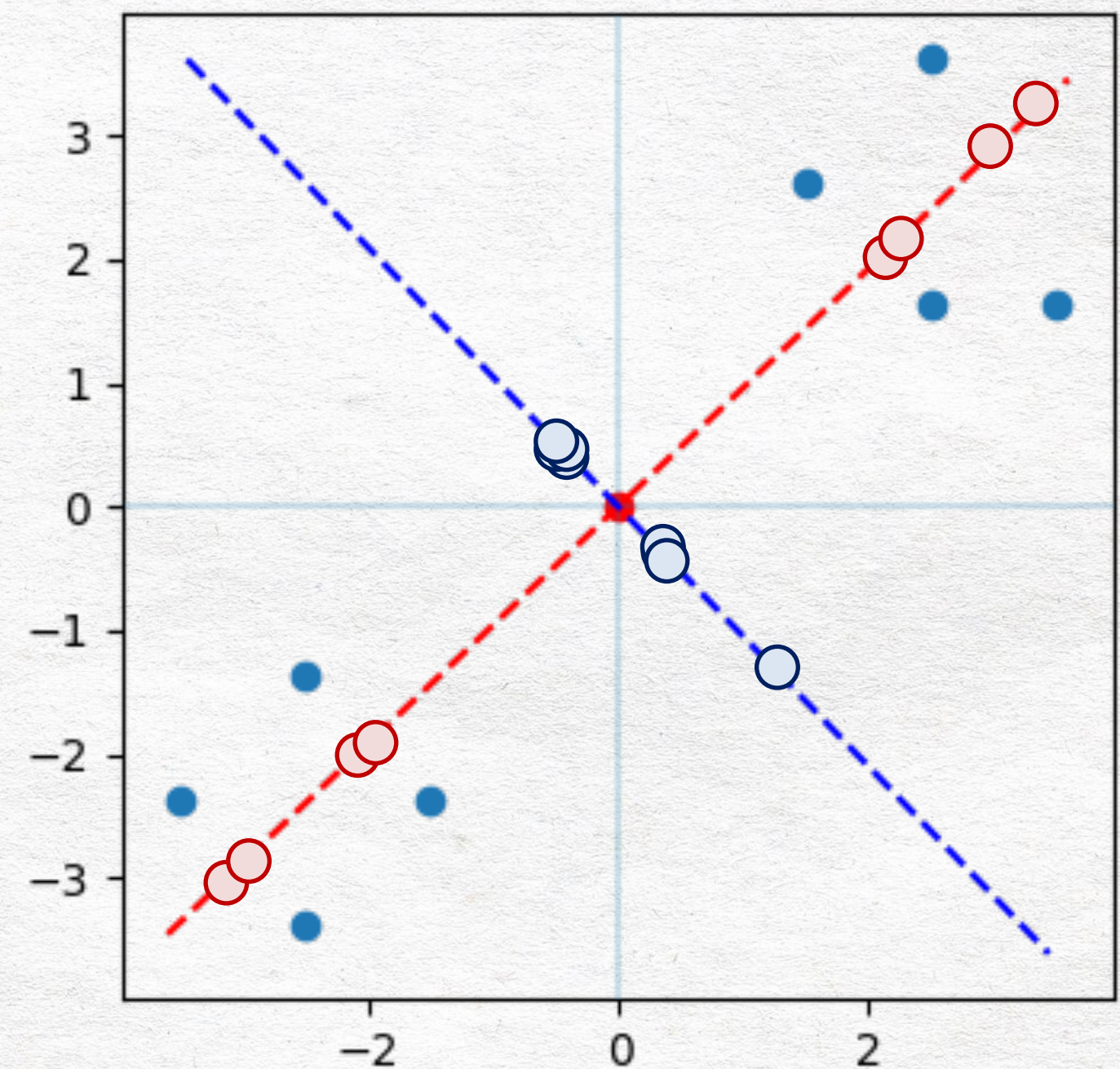
PCA (Principal Component Analysis)

Como vamos fazer o PCA?



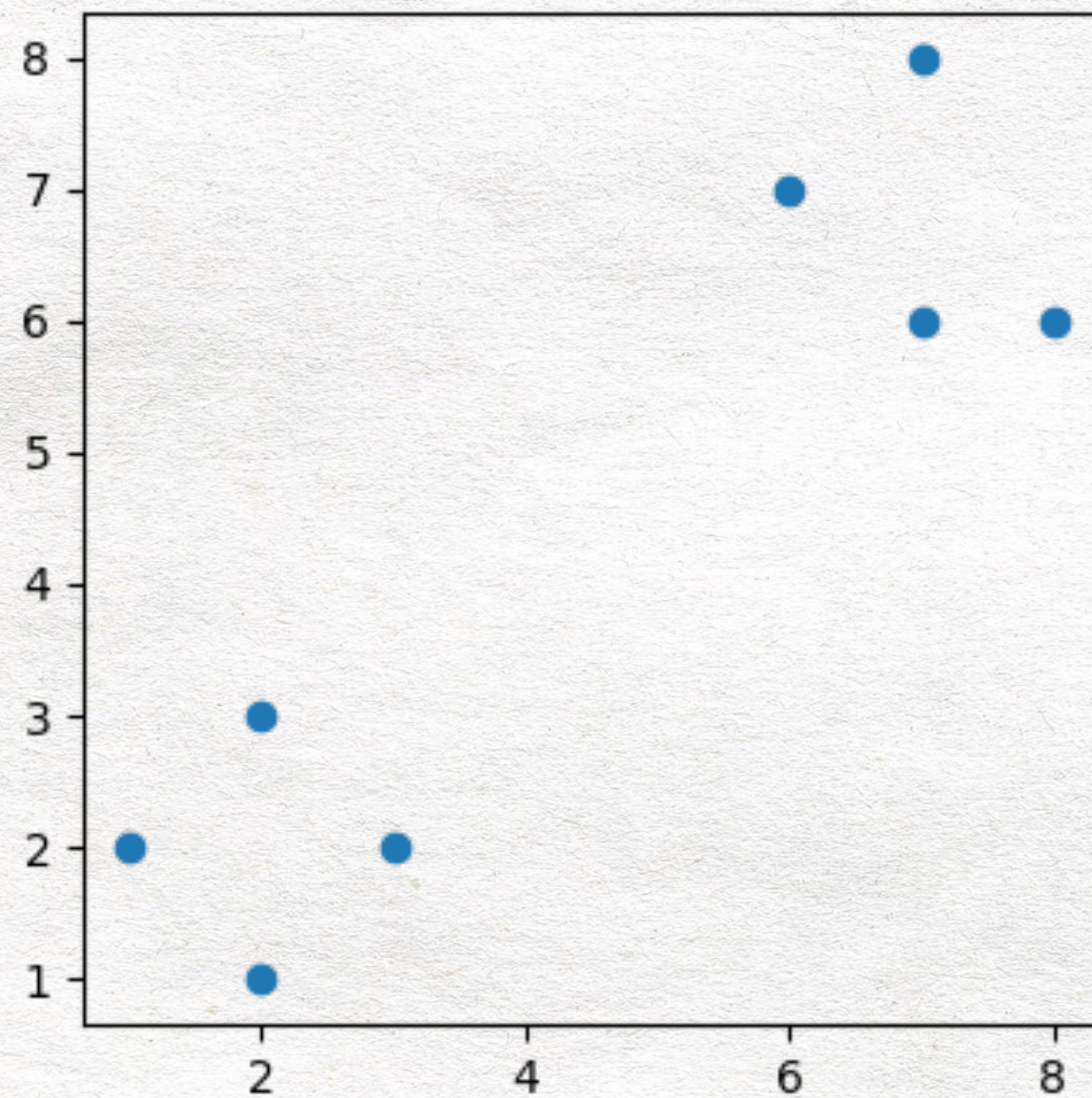
	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

E depois, vamos traçar uma nova reta perpendicular a essa que também minimiza a distância dos pontos a sua projeção



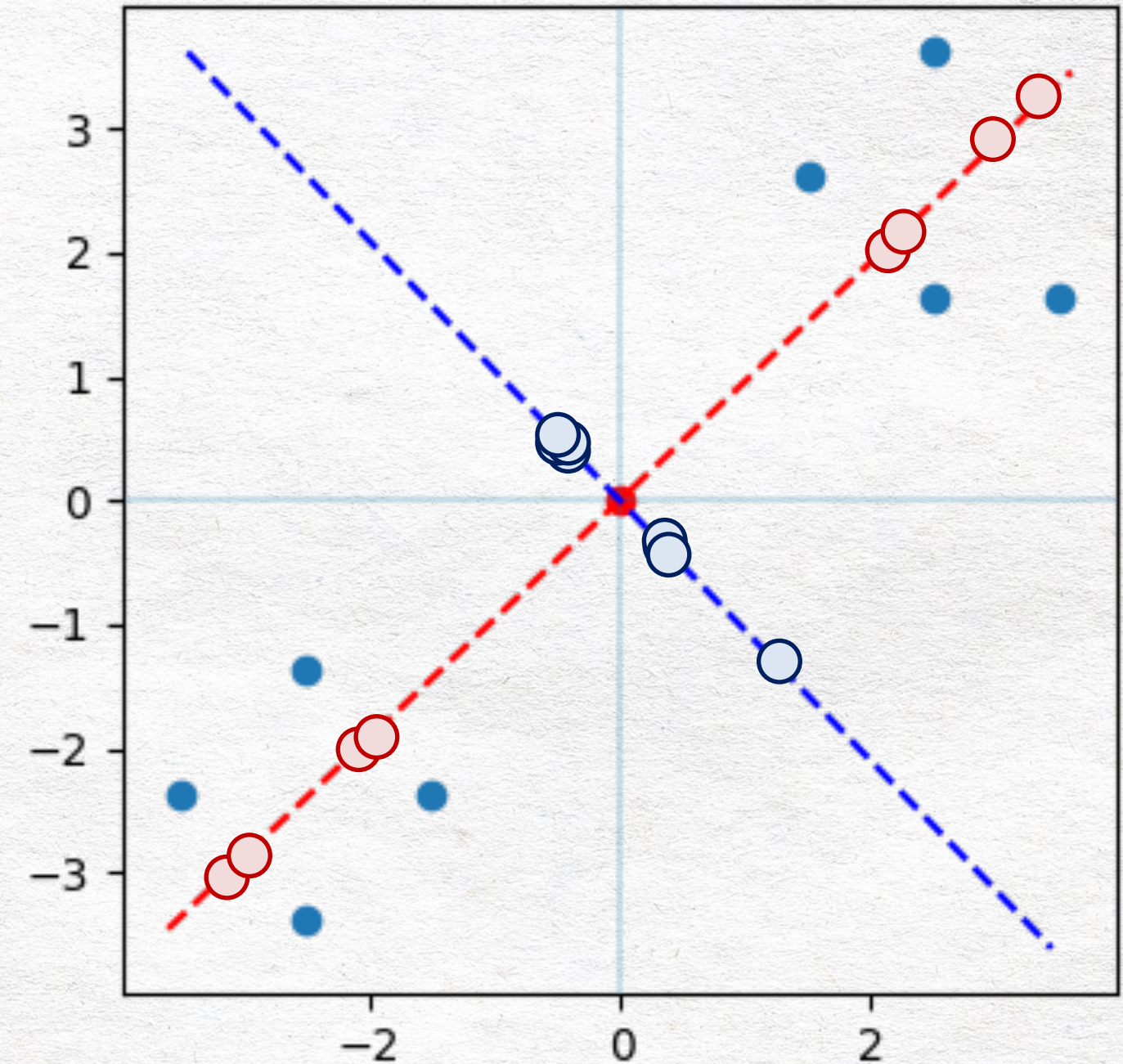
PCA (Principal Component Analysis)

Como vamos fazer o PCA?



	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

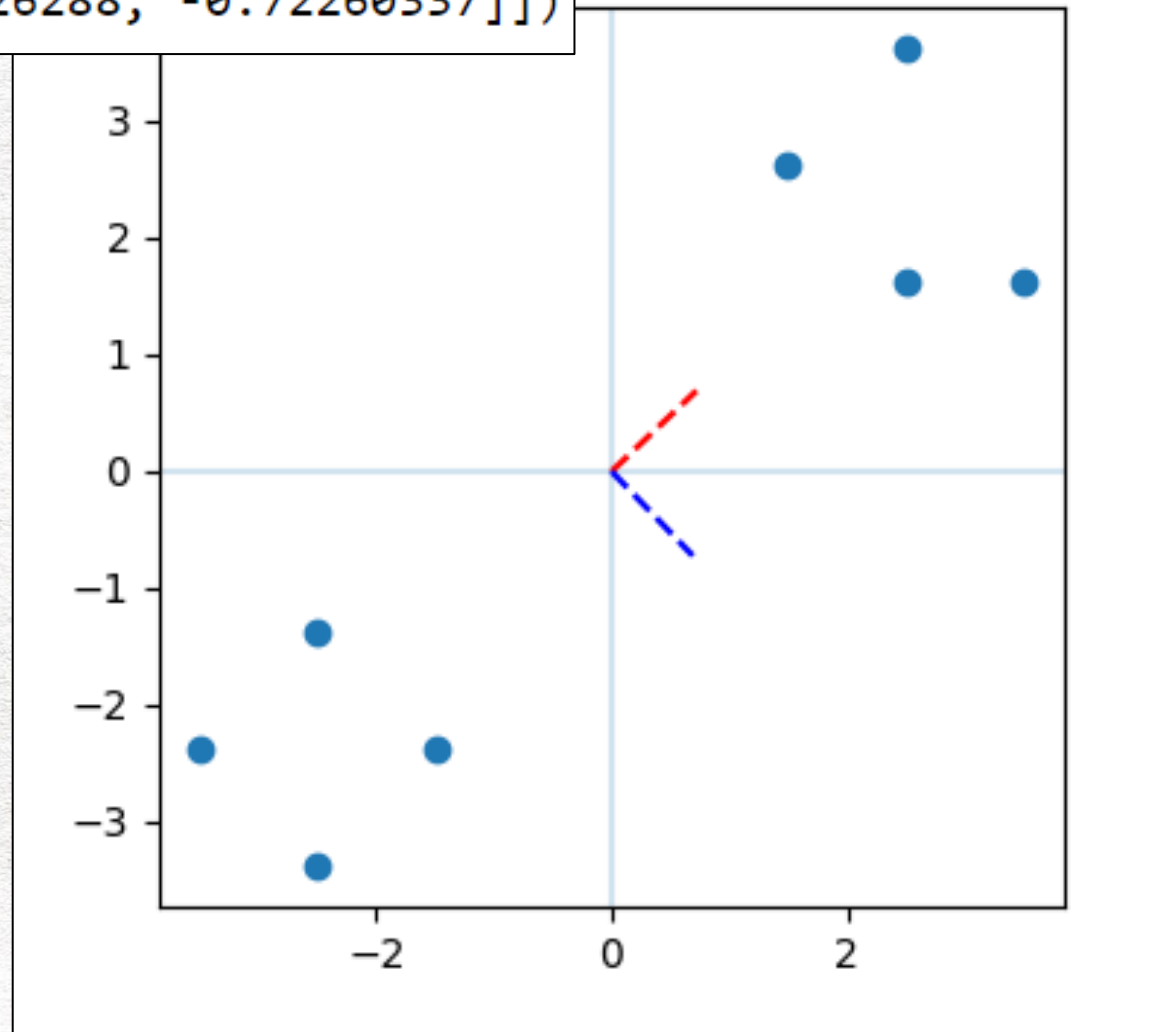
Considerando apenas as retas e as projeções, o primeiro componente deve estar em x e o segundo em y



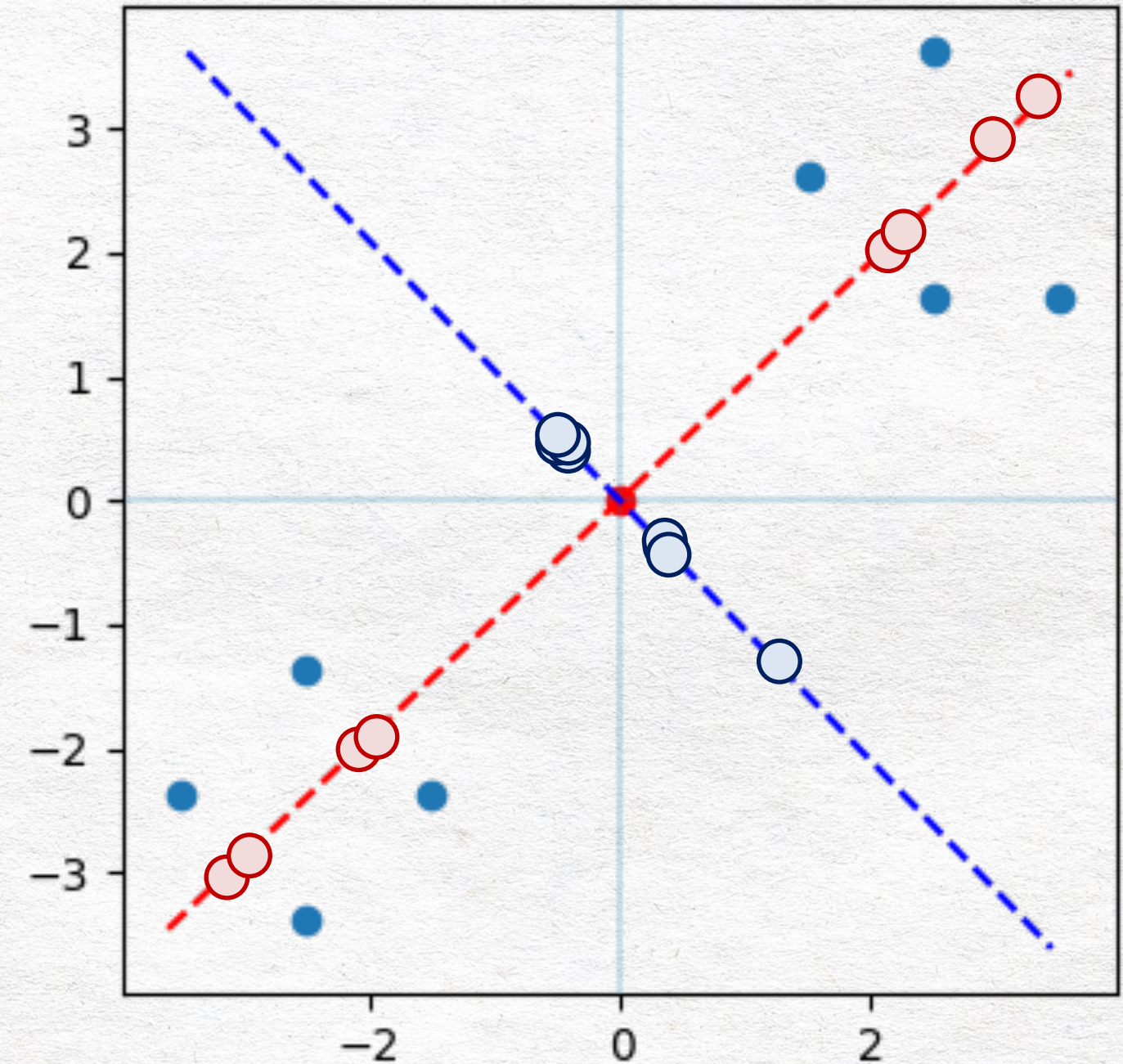
PCA (Principal Component Analysis)

Como vamos fazer o PCA?

```
pca.components_  
array([[ 0.72260337,  0.69126288],  
       [ 0.69126288, -0.72260337]])
```



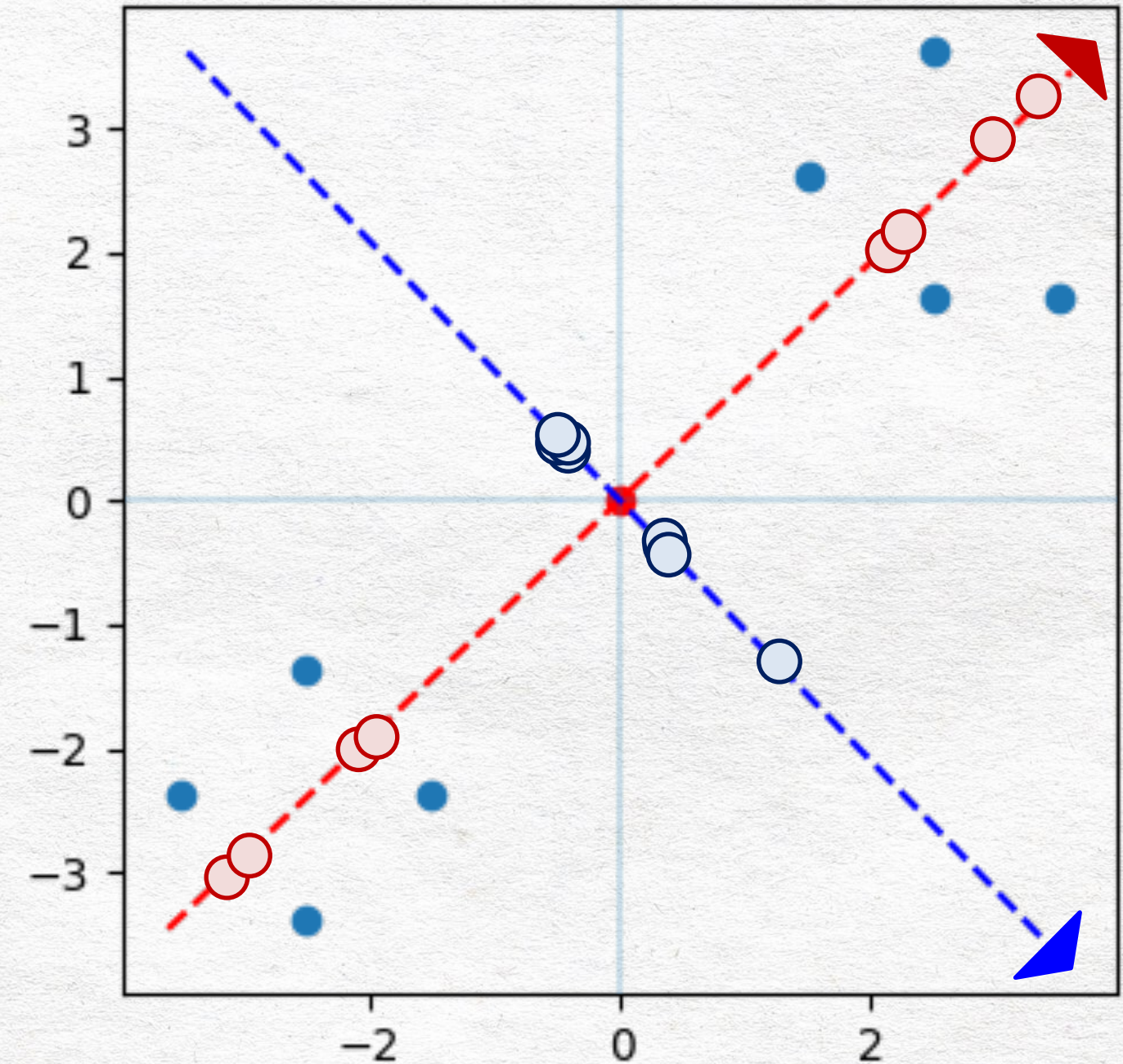
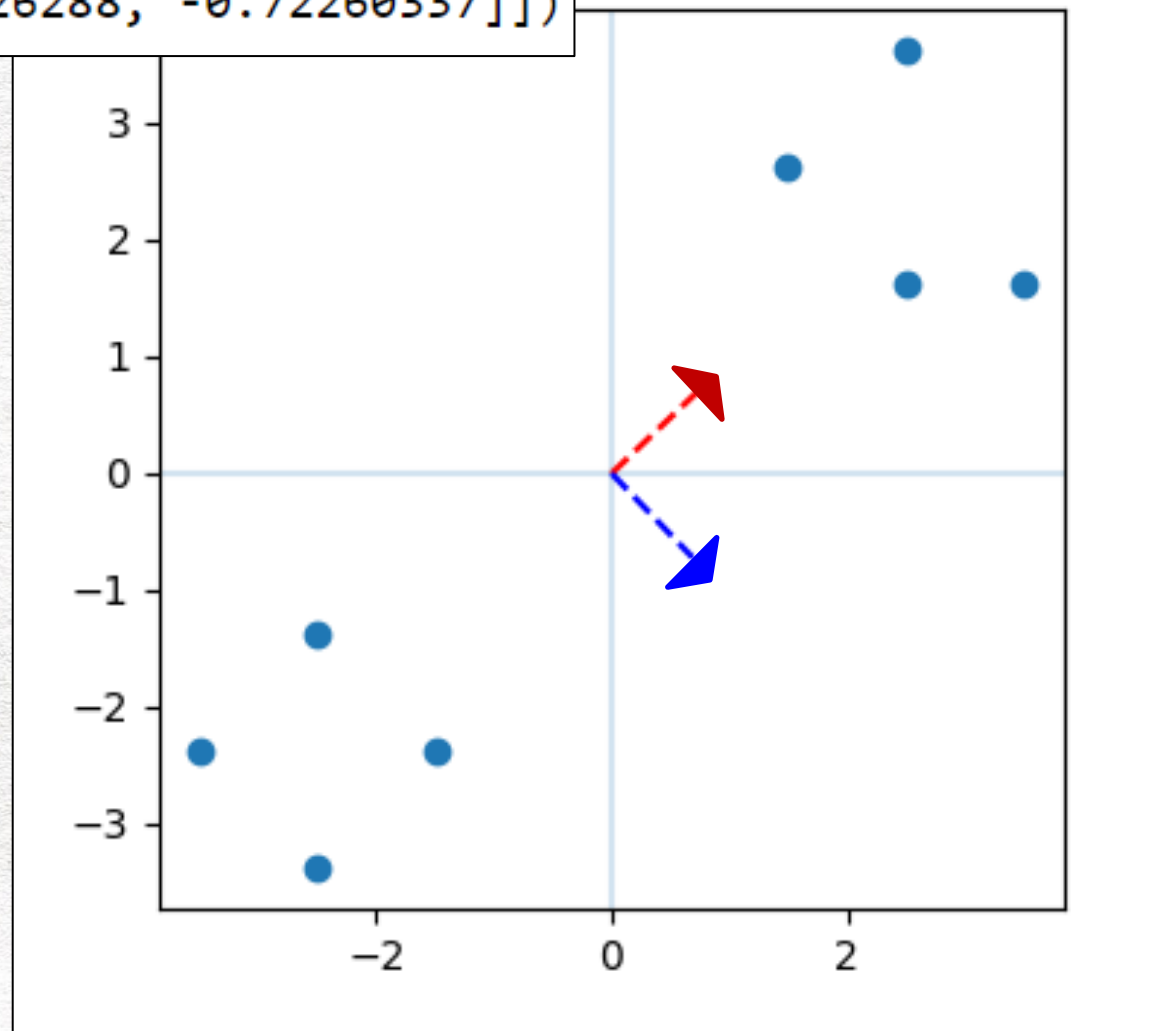
Vamos usar os componentes gerados pelo próprio PCA para indicar como podemos fazer isso



PCA (Principal Component Analysis)

Como vamos fazer o PCA?

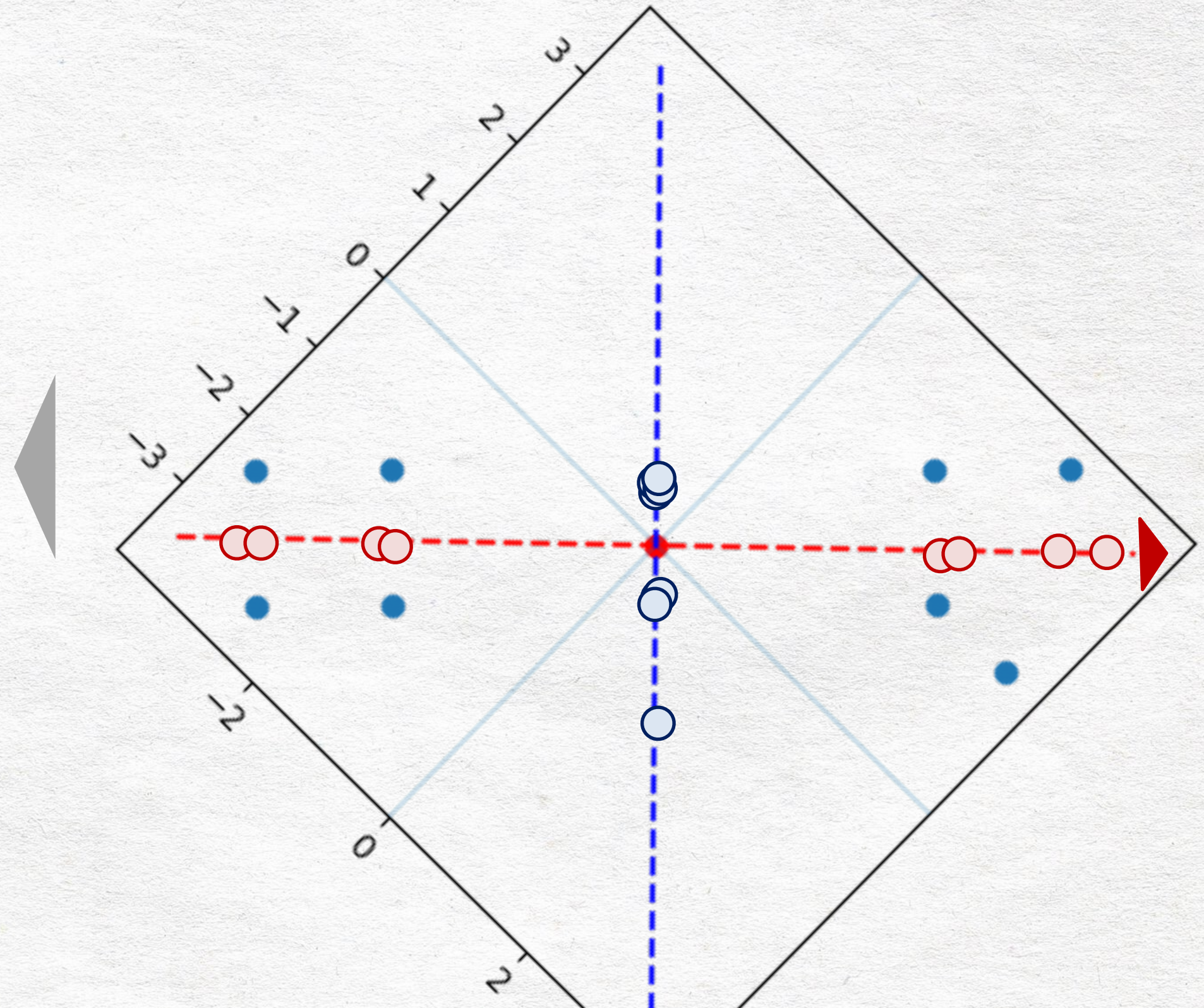
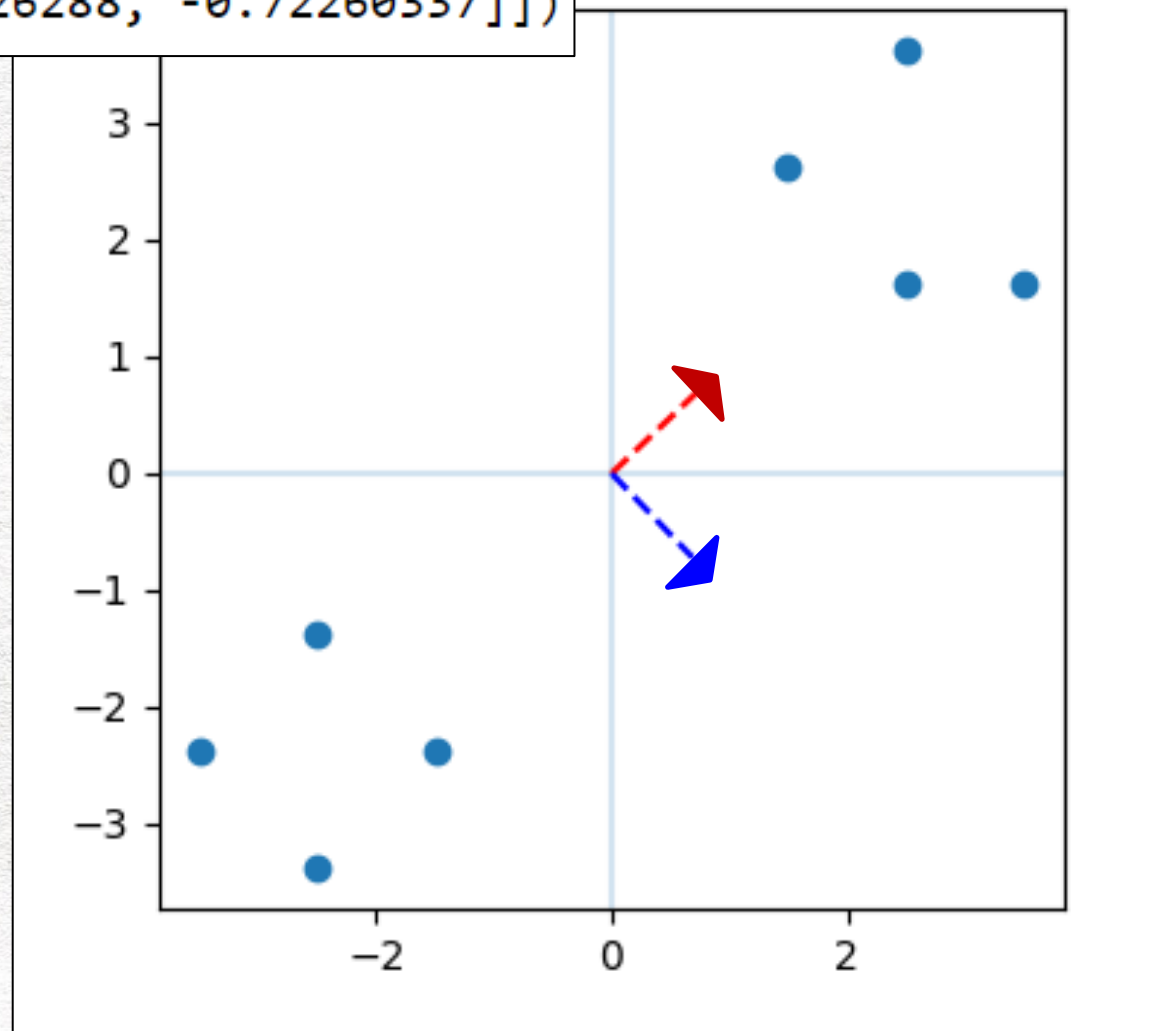
```
pca.components_  
array([[ 0.72260337,  0.69126288],  
       [ 0.69126288, -0.72260337]])
```



PCA (Principal Component Analysis)

Como vamos fazer o PCA?

```
pca.components_  
array([[ 0.72260337,  0.69126288],  
       [ 0.69126288, -0.72260337]])
```

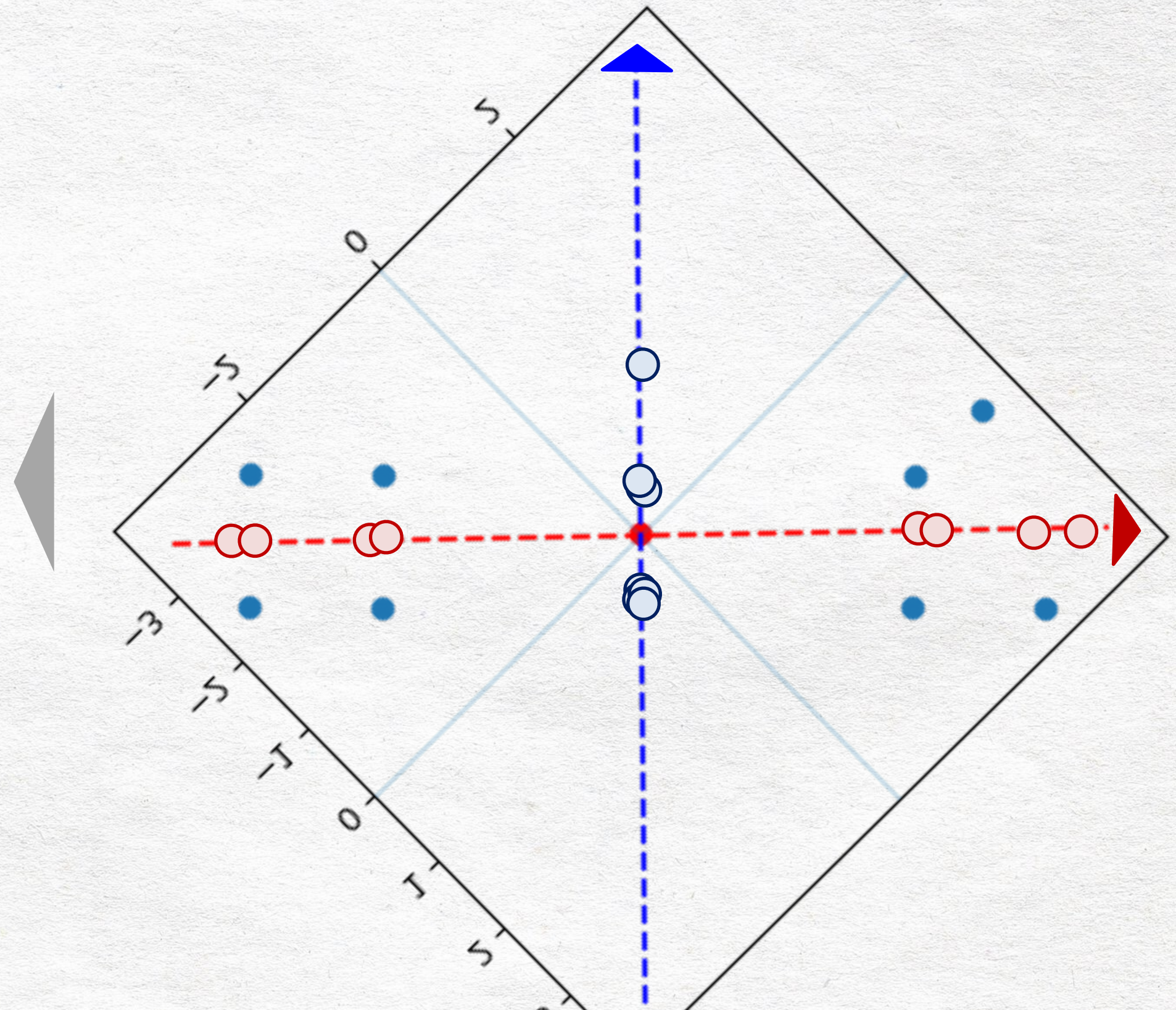
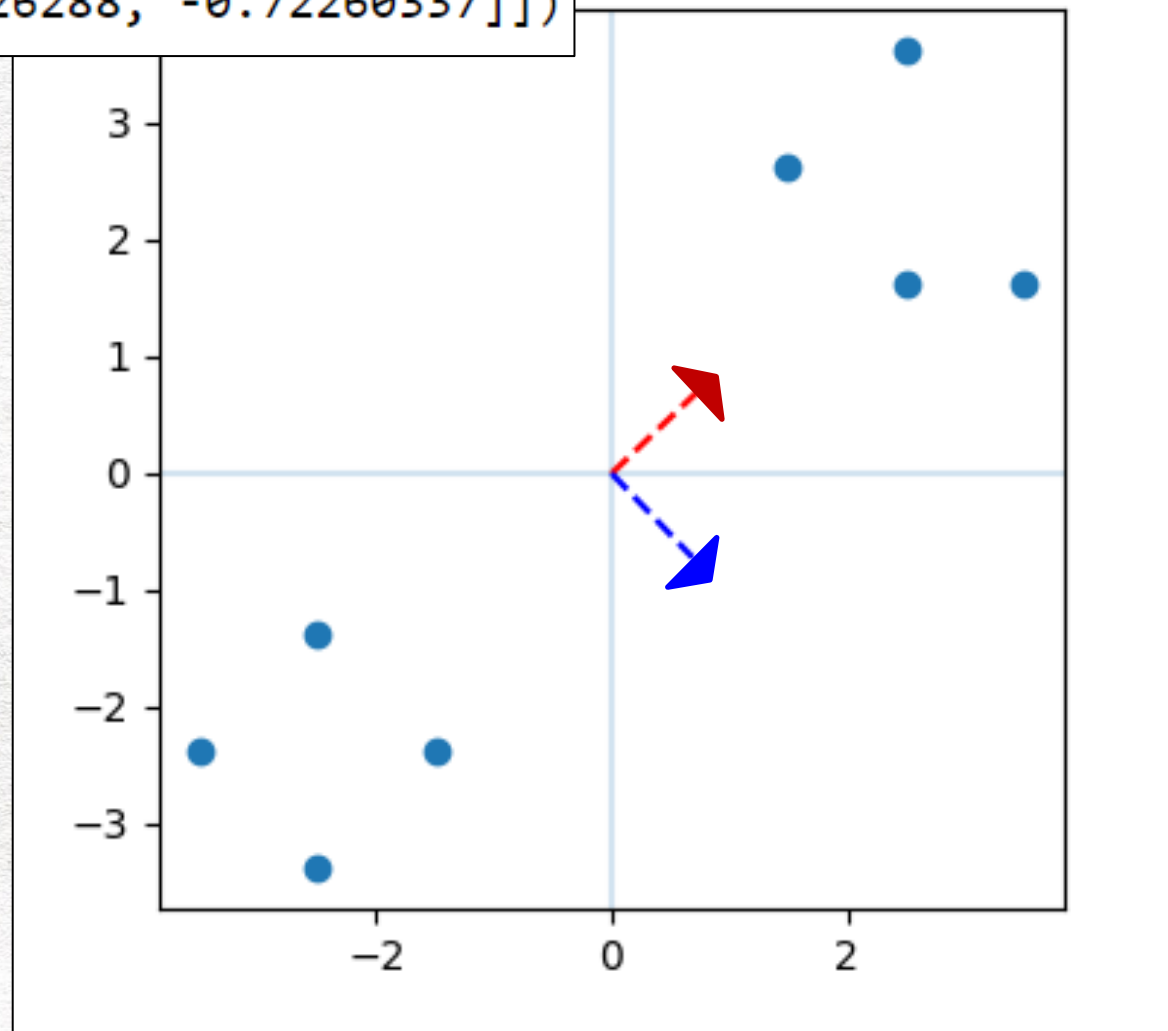


PCA (Principal Component Analysis)

Como vamos fazer o PCA?

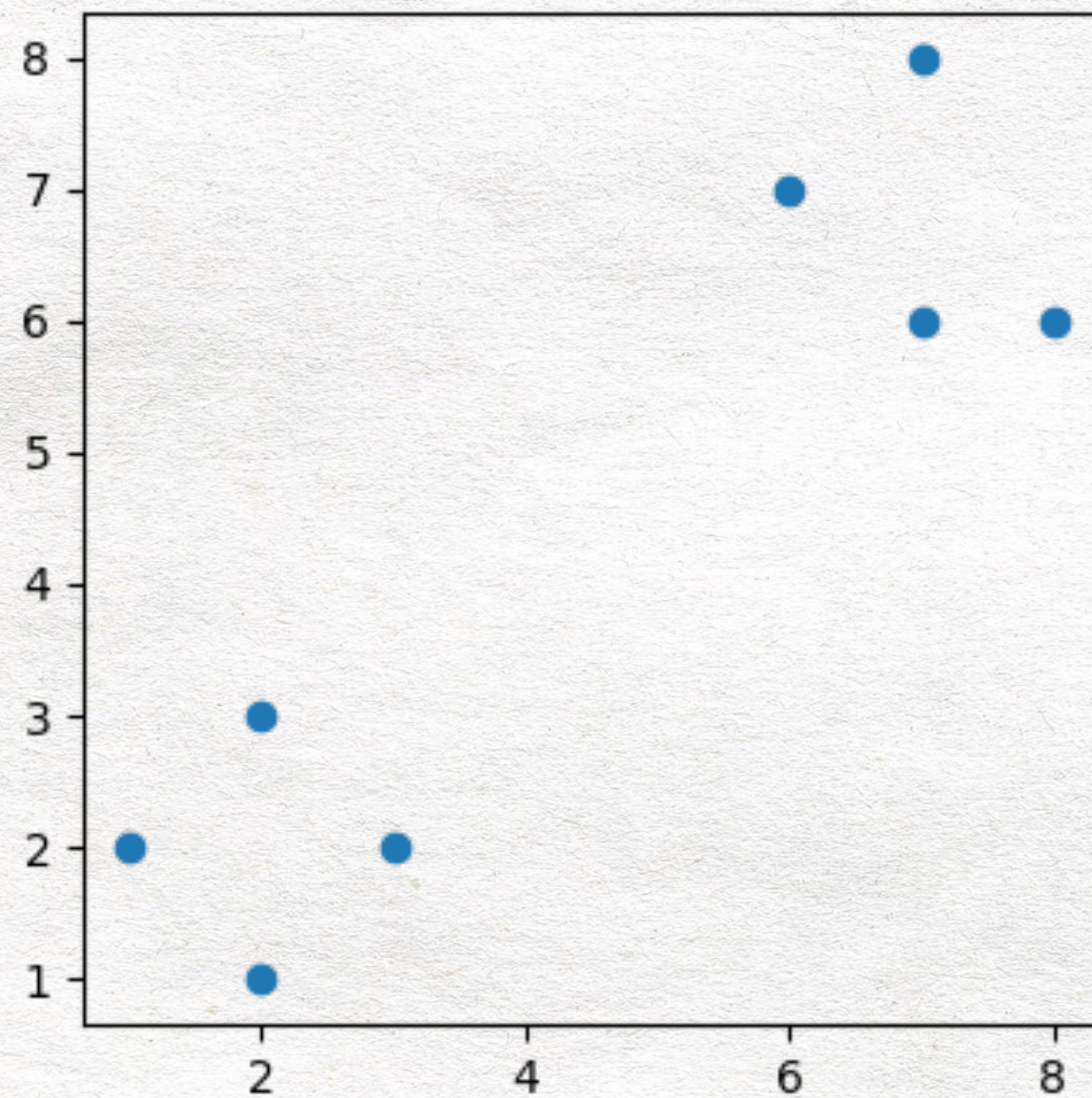
```
pca.components_
```

```
array([[ 0.72260337,  0.69126288],  
       [ 0.69126288, -0.72260337]])
```



PCA (Principal Component Analysis)

Como vamos fazer o PCA?



	x	y
0	1	2
1	2	1
2	3	2
3	2	3
4	7	6
5	8	6
6	6	7
7	7	8

Os pontos de dados serão traçados considerando as duas projeções

